DOCUMENT RESUME

ED 114 432 UD 015 430

AUTHOR Jensen, Arthur R.
TITLE Race and Mental Ability.
PUB DATE Sep 74
NOTE 76p.; Paper presented at a Symposium of the Institute of Biology on "Racial Variation in Man" (Royal Geographical Society, London, England, September 19, 20, 1974)

EDRS PRICE MF-$0.76 HC-$4.43 Plus Postage
DESCRIPTORS Culture Free Tests; Disadvantaged Environment; *Ethnic Groups; Evaluation; Genetics; Heredity; *Intelligence Differences; Minority Groups; Models; *Nature Nurture Controversy; Negro Achievement; Negro Youth; Prenatal Influences; Probability Theory; Psychometrics; *Racial Differences
IDENTIFIERS *Jensen (Arthur)

ABSTRACT

This paper, focusing on differences between Caucasians and Negroes in the United States, summarizes from a "scientific standpoint" the main facts and theoretical issues involved in the study of human racial differences and behaviors. Three principles are considered to govern the orientation of this document: (1) objective research and knowledge are possible and should not be entangled with political and social policy; (2) phenomena are better explained by theory as opposed to ad hoc explanations; and (3) conclusions expressed in terms of statistics and probabilities must be accepted, with their tentative nature kept in mind. Among the topics dealt with are: (1) mutation and drift, migration, and selection processes affecting evolutionary differentiation; (2) formulation of heritability of group differences; (3) psychometric evidence for the author's conclusions; and (4) hypotheses of culture-biased tests and environmental foci. It is reiterated that it is highly probable that genetic as opposed to environmental factors are involved to a substantial degree in the lower average IQ of American Negro groups. (Author/AM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: UD CG
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED114432

RACE AND MENTAL ABILITY

Arthur R. Jensen
University of California, Berkeley

Paper presented at a Symposium of the Institute of Biology on "Racial Variation in Man." Royal Geographical Society, London. September 19 and 20, 1974

To be published in: J. F. Ebling (Ed.) Racial Variation in Man. New York: Academic Press, 1975.

UD 015430

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Race and Mental Ability

Arthur R. Jensen

University of California, Berkeley

Races, both human and infrahuman, are now most generally viewed from a scientific standpoint as breeding populations which, though interfertile, are relatively isolated from one another reproductively, by geography, ecology, or culture, and which differ in the frequencies of various genes. These major subdivisions of a species, called races, are classifications based upon the relative degrees of intragroup similarities and intergroup differences in numerous genetically determined morphological, serological, and biochemical characteristics. These genetic differences are products of the evolutionary process. Some of the many genetically conditioned characteristics in which various human races are known to differ are body size and proportions, hair form and distribution, head shape and facial features, cranial capacity and brain formation, blood groups, number of vertebrae, genitalia, bone density, fingerprints, basic metabolic rate, temperature, heat and cold tolerance, sweating, odor, consistency of ear wax, number of teeth, age of eruption of permanent teeth, fissural patterns on the surfaces of the teeth, length of gestation period, frequency of twins, male-female birth ratio, physical maturity at birth, infant development of alpha brain waves, colorblindness, visual and auditory acuity, ability to taste phenylthiocarbomide, intolerance of milk, galvanic skin response, chronic diseases, susceptibility to infectious diseases, and

pigmentation of the skin, hair, and eyes. Physical differences among some races are obviously extensive and profound.

There are also behavioral differences among races. In infrahuman species, behavioral differences among subspecies (i.e., races) are now generally viewed in an evolutionary sense as being continuous with the physical differences. Ethologists regard behavioral as well as physical traits as being subject to evolutionary change. An animal's behavior can be a more important aspect of its adaptation to the environment than its physical characteristics. and can therefore play an important role in the evolution of the physical structures that mediate behavior, principally the central nervous system.

The biological basis of behavioral differences among human races, on the other hand, has been much more in dispute. There has been the least consensus concerning the nature and causes of racial differences especially in those characteristics which most clearly distinguish Homo sapiens from all other species--a large, highly developed cerebrum and the capacity it affords for complex goal-conscious problem-solving behavior involving planning, reasoning, judgment, imagination, decision, in short, intelligence.

My aim in this paper is to summarize as best I can from a scientific standpoint the main facts and theoretical issues involved in the study of human racial differences in behaviors commonly regarded as indicative of mental ability, without going into the background of socio-political and ideological controversy that continues to surround this topic.

Readers should be told at the outset the three principles that mainly govern my own orientation in this inquiry.

First, I believe that objective research and objective knowledge are possible, and that it is desirable, indeed necessary, to guard the scientific aspects of the matter from entanglement with the political and social policy

aspects. This is not to say that the latter are unimportant, but simply that we should strive as best we can to not let them in any way distort our aim of achieving an objective understanding of racial differences in mental abilities, limited though it may be, considering the intrinsic scientific difficulties.

Second, I emphasize the generally accepted position in science that explanations of phenomena are weak and unsatisfactory to the extent that they are ad hoc, and are more satisfactory to the extent that they are predicted by a more general theory or are consistent with some larger pattern of established systematic knowledge. That theory is best which yields the greatest number of verifiable predictions and the discovery of new phenomena, or can comprehend existing phenemena which previously had only ad hoc explanations. Evolutionary theory, population genetics, the polygenic theory of intelligence, developmental psychology, and psychometrics seem to me to provide the most comprehensive framework for the scientific study of population differences in abilities. It is my belief that explanations of racial differences which do not build upon the theoretical structures of these fields and their associated methodologies are the most likely to be invalid or scientifically unproductive. Whatever theoretical or methodological shortcomings these fields may have at present for the study of racial differences, I know of no better basis for formulating hypotheses and launching investigations.

Third, I believe we must accept the necessarily statistical and probabalistic nature of the evidence and conclusions in many aspects of this research. In comparing populations on psychological traits, we are, of course, dealing with continuous variation involving differences among frequency distributions with marked overlap. Distributions may differ in means, variances, skewness, and

higher moments, and each kind of difference or combination of differences ultimately calls for theoretical explanation and has somewhat different implications. In all of the psychological traits we know of, it is frequently pointed out, variance among population means is much less than variance among individuals within populations. Moreoever, since the causes of population differences in psychological traits are complex, involving many factors which cannot be experimentally controlled in research with human populations, our approach must be largely statistical. Rigorous proof of hypotheses, in the sense of logical necessity or the clear-cut ruling out of all alternative hypotheses by experimental control of variables is not reasonably expected regarding most of the questions of greatest interest. We must make do, at least for the present, with conclusions expressed in terms of probabilities, often rather subjective probabilities at that, based on consistencies among converging lines of evidence and the weight that accrues to hypotheses by virtue of their integration with a larger theoretical framework, as opposed to ad hoc explanations. The tentative nature of conclusions at the growing edge of knowledge should always be kept in mind.

Evolutionary Differentiation

From the viewpoint of evolutionary theory, it is extremely improbable that any genetically conditioned characteristics, physical or behavioral, would have identical distributions of genotypes in all human populations. And the greater the evolutionary separation between any two populations, the greater is the probability of genetic differences in a wide variety of characters. Geographical and cultural isolation of populations over many generations results in cumulative differences in gene pools. The specific evolutionary

processes involved in the genetic differentiation of populations are: gene mutations, random genetic drift, selective migration, and natural selection.

Mutation and Drift. Mutation and genetic drift are random processes occurring at single gene loci, and consequently they are not major causal factors in population differences in polygenic traits, i.e., continuous traits, like height and intelligence, which are determined by a large number of genes. The larger the number of genes involved in a given trait, the less is the probability that random changes, or drift, occurring at individual loci would all happen to act in the same direction to produce large differences between populations.

The theory of genetic drift, however, permits calculations concerning the relative degree of genetic isolation between populations, based on the number of differences that would occur by random genetic drift alone, without considering the greater, systematic and directional differences brought about by selection. On this basis, for example, geneticists have estimated the "divergence times" or extent of genetic separation between the three major races as about 14,000 years between Caucasoid and Mongoloid, 42,000 years between Mongoloid and Negroid, and 46,000 years between Caucasoid and Negroid (Nei & Roychoudhury, 1973). These estimates were based on the observed differences in the frequencies of neutral genes, i.e., genes for which there is no evidence of selection. The divergence time is the time genetic drift by itself would take to make the frequencies of neutral genes differ between the major races as much as they do at present. This means, in other words, that these three major racial groups have been separated long enough and completely enough to permit a purely random genetic drift in gene frequencies equivalent to some 2000 generations of complete separation between the Negroid and the other two

races, and about 700 generations of complete separation between the Caucasoid and the Mongoloid. However, it should be remembered that these differences due to drift would be expected to have only minor explanatory significance for racial differences in polygenic traits, especially traits which have been subject to natural selection.

It is now possible rigorously to measure the evolutionary distances between various species, as has been done with chimpanzees and gorillas, in terms of the degree of similarity of DNA sequence in certain blood proteins, and to measure the evolutionary relatedness of man to the other primates. But as yet this method is not sufficiently developed to delineate the evolutionary distances among human races with any reasonable precision.

Migration. Migration per se is probably not a major factor in producing population differences in polygenic traits. But migration often involves selection, either of the original migrant population or of subsequent generations, since having to cope with the challenges of an alien environment affords new opportunities for selection to alter the gene pool of the migratory groups. For example, migration from a tropical to a temperate clime could involve selection of whatever genes might be involved in the capacity for the planning and foresight needed to survive the long winters. (We know from experimental behavior genetics with animals that the capacity for acquiring almost every behavioral characteristic, including the general capacity for learning, responds to selection.) Also, plagues and famines which often accompanied migrations produced genetic "bottlenecks" in human populations. That is to say, a relatively large population would be reduced for a few generations to a small, highly selected breeding group, with statistically different gene frequencies than the parent population, which then grows again into a larger population. Such "bottlenecks"

can result in marked changes in the gene pool within a relatively short period, depending upon the nature and severity of the selection.

Selection. Natural selection is by far the most probable evolutionary mechanism causing the major differences between human races, especially as regards polygenic traits. When a complex phenotypic characteristic, physical or behavioral, involves the influence of a number of genes, all the genes are selected simultaneously, since selection acts directly on the phenotypes. The rapidity of selection for the relevant genes depends both upon the severity of the selection pressure on the phenotypes and upon the narrow heritability of the characteristic in question, i.e., the proportion of phenotypic variance attributable to additive genotypic variation. Selection, so to speak, tends to use up the additive genetic variance, since it is that part of the individual's genome which is most highly correlated with the phenotype. As selection proceeds, the narrow heritability of the trait decreases; that is, there is less additive genetic variance and an increasing proportion of the genetic variance is attributable to dominance deviation (i.e., interaction or nonadditive effects of alleles at the same loci) and to epistasis (i.e., nonadditive effects of genes at different loci). The presence of nonadditive genetic variation, which can be estimated by the methods of quantitative genetics, therefore, indicates that the trait in question has undergone selection, and if the dominance is for either high or low values of the trait, it means there has been directional selection.

It is highly significant to our inquiry, therefore, that the appropriate quantitative genetic analyses of scores on standard tests of intelligence show some dominance and other nonadditive genetic variance, as much as 10 to 15 percent (Jinks & Fulker, 1970; Jinks & Eaves, 1974). Dominance for high IQ indicates there has been directional selection. Thus, whatever ability is

measured by IQ tests, it shows the "genetic architecture" expected for a fitness character; the IQ apparently reflects some trait of biological relevance in human evolution.

How might this have come about? Cranial capacity, a crude measure of intelligence, is known to have increased markedly over the five million years of human evolution, almost tripling in size from the earliest fossil information of Australopithecus to present day man. The greatest development of the brain was of the neocortex, especially those areas serving speech and manipulation. Tools found with fossile remains indicate that increasing brain size was accompanied by the increasing complexity of tools, and along with the development of complex tools are also found artistic drawings on the walls of caves. In the last one or two million years the strongest selection pressure in man has been for behavioral traits of increasing complexity, accompanied by the increasing size and complexity of the cerebrum. The ethologist Konrad Lorenz (1973) has elaborated upon the thesis that the evolution of the complex functions of the human brain that make possible such intelligent operations as comparing, analyzing, separating, seeing relationships, classifying, counting, abstracting, conceptualizing, recalling, imagining, planning, and the like, came about from selection by environmental demands acting directly upon the behaviors made possible by increasingly complex nervous functions.

It seems highly probable that such powerful selective processes have also operated to some extent differentially upon subgroups within one species that have been genetically isolated for thousands of generations. Therefore, in terms of evolutionary theory, behavioral traits and their genetic and physical underpinnings in the nervous system should be expected, with high probability, to differ among human races. If our psychological measurements

did not reflect such differences, they would seem highly suspect, since, in principle, differences are practically certain to exist.

We can only speculate about which specific selection mechanisms were probably most importantly involved in evolutionary differences in the behavioral capacities now called cognitive ability or intelligence.

Perhaps the most important general factor in selection for brain size and complexity was the presence of other men, making for competition for the means of survival, and selection for increasing ability to cooperate in hunting animals and in conflicts with hostile tribes. The invention of new tools and weapons and the development of skill in their use by other individuals would have conferred differential advantages making for selection. Each new invention in a sense divides the population into those who can and those who cannot learn to master its use, and tends to select in favor of those who can.

Population size is an important factor in the selective advantage of invention. The number of exceptional individuals most likely to make discoveries and inventions is greater, the larger the group. New inventions and novel variations of existing tools and their correlated skills are less likely to arise in the relatively small and culturally isolated groups characteristic of primitive societies. Moreover, when an innovation does occur, and especially if it is a great advance beyond existing knowledge or skill, it may not be perpetuated unless some reasonably substantial number of the group can take it up. Depending upon its degree of novelty and complexity, they would have to be the more exceptionally able individuals, and, given the normal distribution of abilities, more such able individuals would exist in a larger population, so that a new invention of only one exceptional member of the group would take on selective significance for some substantial number of the population.

Inventions and discoveries involving tools, weapons, skills, and knowledge about the environment of adaptive importance, create greater salience of individual differences in abilities which then become important factors in selective and assortative mating. As one moves from relatively primitive to relatively advanced societies, individual differences in cognitive ability become more conspicuous and more consequential in many ways that can affect an individual's fitness in the Darwinian sense. In a number of early human societies mating was a prerogative of the ablest and most esteemed males, each of whom had many females, while many less esteemed males had no mates.

Evolutionary rates for certain traits could differ considerably among groups with different mating customs or different degrees of selective mating for various traits. In considering natural selection for abilities in man, one must consider what proportion of a population is regarded by its members as subnormal or in any way undesirable from the standpoint of selective mating, and this will of course depend to a considerable extent upon the nature and cognitive complexity of the cultural demands made by the society. Even a slight reproductive advantage can have marked genetic consequences on the time scale of human evolution. For example, it can be calculated that a gene that confers a one per cent reproductive advantage in a population will increase in frequency from .01 to over .99 in 1000 generations, assuming that the same degree of advantage is maintained throughout this period.

Increased population size also decreases the degree of inbreeding and gives rise to more new genetic combinations which are grist for selection.

Primitive societies consisted of hunter-gatherers, and for obvious ecological reasons were kept relatively small in numbers. The advent of agriculture permitted population densities a thousand times greater than

those of hunter-gatherers, thus magnifying the selection factors for cognitive abilities associated with a larger population. Also, in terms of abilities for counting, measuring, planning, mastering the environment, and a greater complexity of social, political, and economic organizations, agriculture probably placed a higher premium on intelligence than did hunting and gathering. In fact, civilizations grew up along with the development of agriculture. Various populations of the world differ in thousands of years in the time since they abandoned hunting and gathering for agriculture, and some presently existing groups have never taken up agriculture.

Thus, in general terms, man's evolutionary history and the relative isolation of various populations for thousands of generations would justify the expectation of genetic differences between populations in a host of characteristics including those in which selection pressures have acted differentially upon behavior. These behaviors would be mainly polygenic traits for which population differences are statistical rather than typological. It would seem most improbable that at least some of the genetically conditioned behavioral differences that have come about in the course of evolution would not be among the observable differences between contemporary races. A contrary view would have to argue one of four propositions: either (1) the selection pressures in all long-term isolated populations in the course of human evolution have been identical for all groups for all abilities; or (2) even if there have been different selection pressures for different components of ability, these components would average out to the same value in their combined effects on performance in every population, provided there is equality of opportunity for the development and expression of abilities; or, (3) there is only one general ability that is inherited--a highly plastic capacity for cultural learning which is genetically the same

in all populations and becomes differentiated only through environmental and cultural influences; or, (4) even if there are genetic ability differences between populations, they are so obscured by cultural and environmental factors that there is zero correlation (or even a negative correlation) between the distributions of phenotypes and genotypes. Numbers 1 and 2 have the disadvantage of being extremely improbable. Number 3 is contradicted by the factor analytic and behavior-genetic analysis of mental abilities, which reveal a number of different abilities with relatively independent genetic bases. The fourth point seems more debatable, since it depends so much upon the methods of measuring abilities and the extent of the cultural differences between the groups in question. Modern students of racial differences have seemed most reluctant to point to various aspects of particular cultures as being in themselves indicative of differences in mental abilities. However, John Baker, an eminent biologist who has written recently on the subject of race, notes the fact that racial groups have differed quite markedly in the degree to which they have developed "civilization" (in terms of a list of twenty-one criteria ordinarily regarded as indicative of being civilized) and also the degree to which complex cognitive abilities are manifested or demanded in various societies. The Arunta language of Australian aborigines, for example, conveys only the concrete; abstract concepts are not represented, nor is there any verbal means of numeration beyond "one" and "two" (Baker, 1974, pp. 500-501). Baker notes that these various criteria of cultural and intellectual advancement rank order existing races much as do standard tests of cognitive ability when applied to representative members of these racial groups who have been reared under similar conditions of civilized life. Baker's book is replete with specific factual examples and comparisons of racial groups in terms of these various

criteria. He concludes "the reader will not have overlooked the fact that repeatedly, in each relevant context, the possibility of environmental causes has been reviewed in some detail and rejected as an insufficient explanation of the facts" (p. 533).

The Heritability of Group Differences

The polygenic theory of intelligence attempts to explain a host of phenomena related to individual differences in mental ability, but mainly the form of the distribution of intelligence in the population and the degree of resemblance, or correlation, of mental test scores between various kinships. It is a fact that the degree of correlation between individuals' intelligence test scores increases systematically with the closeness of their genetic kinship, from identical twins, at the one extreme, to unrelated persons, at the other. The polygenic theory, based on principles of Mendelian genetics and its elaboration, called biometrical genetics, for dealing with polygenic traits, yields predictions of these various kinship correlations, and the fit of the model to the empirically obtained correlations is remarkably good (Burt & Howard, 1956; Erlenmeyer-Kimling & Jarvik, 1963; Jinks & Fulker, 1970). (The quantitative-genetic aspects of the model have been most clearly explicated by Burt [1971]. The current status of the polygenic theory from the viewpoint of the philosophy of science has been critically examined by Urbach [1974].) The pattern of kinship correlations for intelligence test scores closely resembles that for other polygenic traits such as physical stature. The fit is not perfect, however, because of errors of measurement and the influence of environmental factors on intelligence.

The methodology of quantitative genetics, which is derived from the general polygenic model, makes possible the estimation of the proportions

of phenotypic variance in the trait attributable to genetic and to environmental factors. Within this framework a large number of studies have yielded estimates of the proportion of genetic variance (called heritability or $\underline{h}^2$) in intelligence test scores in the range from .60 to .90, with most of the estimates between .70 and .80. (This is equivalent to a correlation of about .80 to .90 between phenotypes and genotypes.) Thus, the remaining 20 percent of the variance would be attributable to environmental effects, both prenatal and postnatal, and to errors of measurement, i.e., the imperfect reliability of the tests.

The polygenic theory of intelligence is based on three assumptions: (1) there is a general factor of mental ability which is manifested to some degree in all complex mental tasks requiring choice, judgment, abstraction, grasping relationships, etc.; (2) this general ability can be measured in individuals more or less reliably by standard intelligence tests and can be distinguished from other kinds of abilities, acquired skills, and sensory-motor capacities; and (3) individual variation in this ability is the result of a number of genes (probably not fewer than 20 nor more than 100), each having small, similar, independent additive effects, plus a smaller number of genes having interactive effects (i.e., dominance and epistasis). (In addtion, but not an intrinsic aspect of the polygenic theory, there are mutant or defective genes, often called "major genes," with a very low frequency in the population, the single occurrence of which completely overrides the normal polygenic determinants of intellignece to produce one of the severe clinical forms of mental deficiency.) In accord with the Mendelian principles of random segregation and recombination of genes, this polygenic model accounts for the normal or Gaussian distribution of intelligence. (The slight but significant empirical deviations from normality are accounted for by major

genes and brain damage due to trauma, disease, etc., as well as differential degrees of assortative mating of individuals scoring in the upper and lower halves of the distribution, and covariance of genetic and environmental effects.)

The polygenic theory of intelligence is one of the few well developed and well substantiated models in psychology. There simply is no competing model that comprehends the relevant facts. Environmentalists who oppose a genetic theory of individual differences in ability have proposed no alternative theory to account for all the facts predicted by the polygenic model. They offer only ad hoc criticisms of specific empirical tests of the polygenic model.

Thus, the polygenic theory, scientifically speaking, gives a quite good account of individual variation in intelligence. In principle, at least, it is also applicable to group differences in ability, which are viewed as qualitatively the same as individual differences. The gene pools of relatively isolated populations are hypothesized to differ in the frequencies of the genes involved in abilities. But the relevant genes are the same in all populations, so their differences are quantitative, not qualitative. The polygenic theory itself is completely agnostic as to the direction and magnitude of the genetic difference between any two specific populations. In this respect the polygenic theory contrasts markedly with the environmentalist view, which maintains that there are no genetic differences in mental abilities, or at least in general intelligence, among any human populations.

It may seem surprising that, in practice, the polygenic theory yields few predictions concerning differences between particular races which are testable by means of any presently available evidence. One type of prediction concerns the intelligence of racial hybrids. The polygenic theory

predicts the mean scores (say, IQ) of the hybrid offspring to be approximately intermediate between the means of the two different racial populations. The only studies of this type reported in the literature are of white and Negro crosses. These studies are generally unsatisfactory, as I have pointed out in detail elsewhere (Jensen, 1973, pp. 219-230), since there is reason to believe that persons entering into interracial marriages are probably not representative of their populations in intelligence. Most studies of the intelligence of racial hybrids are not based on known pedigrees, but on the selection of hybrid subjects solely on the basis of their physical appearance being more or less intermediate between Caucasoid and Negroid in such characteristics as skin color, nasal width, and interpupillary distance. The majority of such studies are in accord with the genetic prediction, i.e., the intermediate group in appearance also usually stands between the more "pure" appearing racial groups in mental test scores. Also, in 12 out of 18 studies of American Negroes with some Caucasian admixture, there was a significant positive correlation between skin color (lightness) and IQ. Although these studies leave little doubt of a relationship between skin color (and other racial characteristics) and IQ, they are a weak test of the genetic theory, since the same correlation could result from cross assortative mating for skin color and IQ within the Negro population without any necessary implications concerning the direction or magnitude of a possible genetic difference in IQ between the Negro and white populations.

To overcome this problem, it has been proposed to use socially invisible genetic polymorphisms, which differ in known frequencies in West Africans and Europeans. These blood polymorphisms could be used as an index of racial admixture which would be correlated with IQ independently of visible racial characteristics such as skin color. Aside from the technical difficulties in such

research, which I imagine are surmountable, there seems a serious conceptual problem with this approach due to the fact that little is known about the selection that entered into interracial matings during the period of the greatest gene flow from the white to the American Negro gene pool, which occurred during slavery. The Negro population of the United States now has an average Caucasian admixture of about 25 percent. We do not know how representative of the white population in intelligence were those individuals (practically all males) who practiced miscegenation. If they were predominantly from the lower half of the white IQ distribution and their mates predominantly from the upper half of the Negro IQ distribution, the genetic consequences of hybridization on the IQ distribution of subsequent generations of American Negroes could be negligible or undetectable by any presently available methods of genetic analysis.

Another prediction from genetic theory involves the phenomenon of "regression to the mean." The offspring of exceptional parents, i.e., those who deviate above or below the population mean, average some value more or less intermediate between the parental value and the mean of the population. Regression is also observed in the case of siblings. Sibling regression is less likely to be contaminated by environmental effects than parent-child regression, since not infrequently a parent has grown up in a quite different environment than is afforded to his or her own children. Siblings reared together generally share a more common environment. The regression is strictly predictable from the polygenic model, but the degree to which the empirical findings approximate the prediction depends upon the heritability, $\underline{h}^2$, of the phenotypic measurements. The complement of the heritability, $1-\underline{h}^2$, consisting of environmental and error variance, can be regarded as "noise" obscuring the prediction. For traits of high heritability, such as height and intelligence,

the predictions are confirmed fairly precisely. Since the theoretical genetic correlation between siblings is 1/2 under random mating and slightly higher (about 0.55) under the degree of assortative mating generally found for intelligence, one should expect, on average, that the IQ of a given child's sibling would be just about half-way between the given child and the population mean. Thus, it is predicted from the genetic model, for example, that the siblings of white and Negro children who are perfectly matched for some given IQ would, on the average, have different IQs, since the Negro sibs regress toward the Negro population mean and the white sibs regress toward the white population mean. If the two populations differ by about one standard deviation (or 15 IQ points) the two groups of siblings of the IQ-matched Negro and white groups should differ half a standard deviation. If the two IQ-matched racial groups both have an average IQ of 120, for example, the average IQ of the Negro sibs will be 1/2(120-85)+85 = 102.5 and the IQ of the white sibs will be 1/2(120-100)+100 = 110.

This prediction was borne out in a study of all the Negro and white siblings in the elementary schools (ages 5 to 12) in a California school district (Jensen, 1973, pp. 117-119). The siblings of both white and Negro children were found to regress a constant fraction, about one-half, to their respective population means and not to the mean of the combined populations. This holds throughout the IQ range from about 50 to 150; the regression line, for both Negroes and whites, is linear throughout that range. Thus, this is a successful prediction from the genetic model. But it cannot be regarded as a proof of a genetic difference between the two populations, since the lower population mean of the Negro group, it could be claimed, is a result of a uniform environmental disadvantage or test bias in the Negro population. Thus, all that the sibling regression demonstrates rigorously is that the correlation

between sibs is about the same in the white and Negro populations. A strictly environmental explanation of the mean population difference is not ruled out by this evidence. But an environmental explanation of it is ad hoc, unlike the genetic explanation, which is derivable from a pre-existing polygenic model. The polygenic theory would be in serious trouble if the prediction were not borne out. But there is no environmental theory that would have predicted the quantitative aspects of these results or the linearity of sibling regression throughout the normal range of IQs. In an ad hoc environmental account of the results, it would have to be regarded as a remarkable coincidence that environmental factors would so closely produce the same quantitative effects as are predicted by the genetic model.

Essentially, the reason that the regression phenomenon by itself does not prove genetic difference between populations is that even if one grants the same degree of heritability of a trait within each of two populations, and even if the heritability is very high, it cannot be inferred with certainty that the difference between the populations has a genetic component. It could be all environmental, or all genetic, or anything in between.

It is generally agreed that heritability within-groups, h^2_W, has no logically necessary implication for heritability between groups, h^2_B. This does not imply, however, that there may not be probabalistic implications of h^2_W for h^2_B or that there is no theoretical connection whatsoever between h^2_W and h^2_B, given knowledge of certain other parameters.

Generally, for highly heritable characteristics within groups, phenotypic mean differences between groups also show a heritable component, even when there are obvious environmental differences between the groups. Often there is a positive correlation between genotypes and the environmental factors

most relevant to the characteristic, e.g. skin pigmentation and amount of exposure to ultraviolet radiation.

Instances are rare where the direction of genotypic means is the opposite to that of the phenotypic means; more often phenotypic and genotypic means are positively correlated. If within-group heritability is high (i.e., greater than 0.5), one must hypothesize a larger environmental difference than a genetic difference to explain a phenotypic difference between group means, unless one also posits an additional hypothesis that the mean difference between groups is due to environmental factors which are not the same as those responsible for environmental variance within the groups.

A reasonable presumption (though certainly not proof) of genetic group differences seems to be related to the magnitude of the group difference and the heritability of the trait in question, as seen in the fact that few persons believe that the average difference in stature between Pygmies and Watusis is not largely genetic, despite their very different habitats, diets, and customs. The fact that the group mean difference is large (relative to the standard deviation within groups) and involves a trait of very high heritability, makes it seem reasonable to believe that the group difference is largely genetic. (I know of no other evidence that it is genetic.) The same kind of "reasonable hypothesis" must also apply to other characteristics including behavioral traits, in which there are substantial phenotypic differences and substantial heritability within groups, although, of course, the degree of plausibility will depend upon the magnitudes of the group difference and of the within-groups heritability of the trait in question, as well as upon other factors such as the nature and extent of environmental differences, if these are known and their causal relationship to the trait in question is established.

Formulation of Between-Groups Heritability as a Function of Within-Groups Heritability.

The geneticist Jay L. Lush (1968) proposed the following formula of the relationship of between-groups heritability, h_B^2 (i.e., the genetic fraction of the variance among the phenotypic group means) and the heritability in the whole population (i.e., the combined groups):

$$h_B^2 = h^2 \left[\frac{1 + (n-1)r}{1 + (n-1)t} \right] \qquad (1)$$

where h^2 is the narrow heritability in the whole population

n is the sample size

r is the intraclass correlation among the genic values (for the particular character in question) of members of the same group.

t is the intraclass correlation among the phenotypic values of the same group.

When n becomes large,

$$h_B^2 \cong h^2 (r/t). \qquad (2)$$

The heritability within groups, h_W^2, can be expressed as:

$$h_W^2 = h^2 \frac{(1-r)}{(1-t)} . \qquad (3)$$

From Equations 2 and 3, the geneticist De Fries (1972) derived the following formula for the heritability between groups:

$$h_B^2 \approx \frac{(1-t)r}{(1-r)t} . \qquad (4)$$

If there is a positive correlation between heredity and environment, this expression underestimates the heritability of the group difference. If the correlation between heredity and environment is negative, h_B^2 is overestimated by the formula. The relationship of between-group to within-group heritability for two groups with equal variance, normal distributions of the trait, and a mean difference of one standard deviation, can be shown graphically as in Figure 1.

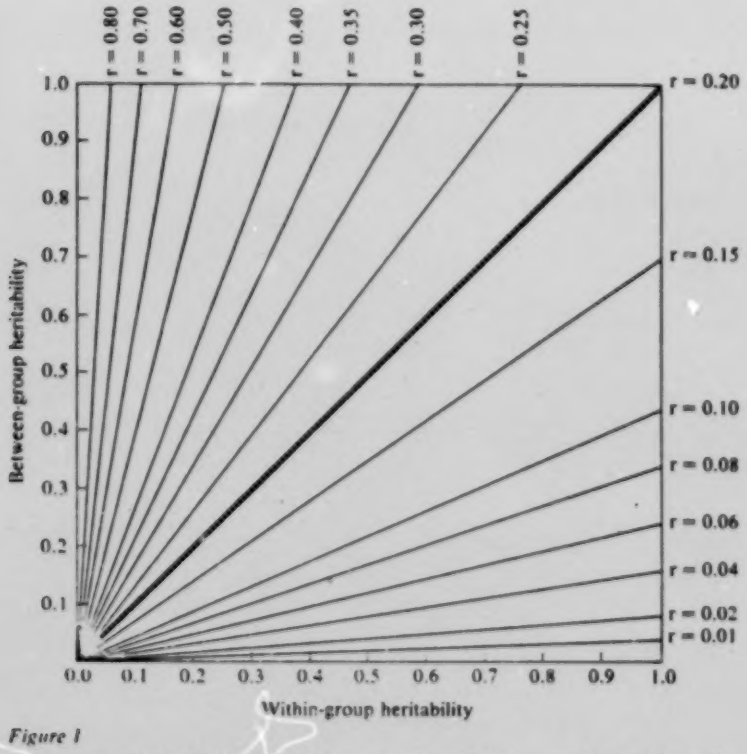

Figure 1

Between-group heritability, expressed as a function of within-group heritability and the genetic correlation of members of the same group (r). See text for explanation.

(From McClearn & De Fries, 1973, p. 300.)

The formula is obviously only of theoretical interest, since we lack information on one of the parameters, r, the intraclass genic correlation for the trait in question. Thus the formula gets us nowhere, unless, of course, one wishes to speculate concerning the probable value of r. But this is the very point in question. If the groups do not differ at all genetically, r will be zero and h_B^2 will be zero. For groups whose means differ by one standard deviation, the phenotypic intraclass correlation, t, is 0.20. (The intraclass correlation t = .20 is most easily obtained from a one-way analysis of variance which partitions the total variance (say, of IQ) between-groups and within-groups. If the group means differ by 15 IQ points and the σ within each group is 15 IQ points, then the between-groups variance σ_B^2 will be $(15/2)^2 = 56.25$, and the within-groups variance σ_W^2 will be $15^2 = 225$. The intraclass correlation is $t = \sigma_B^2/(\sigma_B^2 + \sigma_W^2) = 56.25/(225 + 56.25) = 0.20$.)

The genic intraclass correlation r, however, is unknown. Unless one assumes that all the genetic difference between groups in the trait of interest is purely a result of random genetic drift (which affects all gene loci equally, on the average), there is no way I know of to estimate r for any particular polygenic trait. And the traits in which we are most interested psychologically probably do not involve exclusively neutral genes. If they did involve only neutral genes and the trait were highly polygenic, then there would be no reason to expect any appreciable systematic genetic difference between large population groups. The size of r will of course differ for various traits which have been subjected to different selection pressures over many generations. Thus it is pointless to try to estimate r for one characteristic and expect it to be generalizable to others.[1] The De Fries formula therefore is useless

empirically. Those who believe there are no genetic differences will say $\underline{r} = 0$. If one makes the unwarranted assumption that genetic group differences are not confounded with environmental differences, then it might be said that $\underline{r} = \underline{s}/\underline{h}^2$ (where $\underline{h}^2$ is the heritability in the whole population). And if one makes the assumption that the between-groups environmental effects are of the same nature as within-groups environmental effects, one could say that $\underline{r} = \underline{h}^2\underline{t}$. But without making that assumption, which is crucial to the whole argument, we cannot know $\underline{h}^2$ in the combined populations, either, since this $\underline{h}^2$ is a function of $\underline{h}^2_W$ and $\underline{h}^2_B$, and it is $\underline{h}^2_B$ that we can't determine for lack of knowing $\underline{r}$. Because of this lack, we must conclude that, at present attempts to infer the magnitude of heritability between groups is a blind alley.

Within- and Between-Groups Environmental Variance. A knowledge of the heritability of intelligence within each racial group places some constraints on the magnitude of the mean difference between groups that can be accounted for in terms of all the environmental factors that contribute to variance within groups. The argument can be expressed most clearly in a series of points, using the well established White-Negro IQ difference (in the United States) of one standard deviation, as an example.

1. If the heritability $\underline{h}^2$ of IQ is 0.7 to 0.8 in the white population (which is the best estimate we have from consideration of the total evidence), then the proportion of IQ variance attributable to environmental factors is $1-\underline{h}^2$ or 0.2 to 0.3. The standard deviation of the total environmental component

of IQ thus can be calculated to be about 6 to 8 IQ points (i.e. if σ of IQ is 15, the variance is 15^2 and the proportion of environmental variance would be $0.2 \times 15^2 = 45$, so the standard deviation would be $\sqrt{45} = 6.7$).

2. If one assumes similar heritability of IQ in the Negro population, the standard deviation of the environmental component of IQ is about the same as in the white (i.e. item 1 above). (The evidence for IQ heritability in Negro populations, though not strong, does not suggest that h^2 differs appreciably from the estimate in white populations.) The existing correlations for twins and for siblings are highly similar for Negroes and whites. This does not prove that the heritability is the same in both groups, but it makes it the most likely hypothesis.

3. If white and Negro populations differ on average, by some 15 to 20 IQ units, as the preponderance of the evidence indicates, then, given points 1 and 2 above, if it is hypothesized that all of this difference is environmental, it must be concluded that the groups differ by about two to three standard deviations in all the nongenetic sources of variance that make for IQ differences within the groups. Few would claim that the micro-environmental factors that constitute the within-families variance (e.g., birth order) should be included among the causes of the average difference between populations. It is the sources of between-families variance, i.e. the kinds of environmental factors affecting all members of a family, that contribute to social class and racial group differences in IQ. The between-families environmental variance is about one-half to two-thirds of the total environmental variance within racial groups. This means that Negroes and whites, on average, must differ by some 3 to 4 standard deviations in such environmental influences if the standard 15 to 20 points IQ difference is to be explained entirely in these terms.

4. A variety of socioeconomic indices, singly and in combination, indicate

that the average White-Negro difference in this respect is about one standard deviation or less--far from the 3 or 4 standard deviations of environmental difference that must be postulated by a strictly environmental hypothesis of the White-Negro IQ difference. In terms of these measurable (and potentially manipulable) kinds of environmental factors, studies of adopted children suggest that moving one standard deviation up or down on the environmental scale pushes the child's IQ up or down some 6 or 7 points. Hence these kinds of environmental factors can account for only about one-third to one-half of the White-Negro IQ difference.

As one example, we can look at a study by sociologist Jane Mercer (1973), which includes an exceptionally detailed rating of environments of large samples of white and Negro school children.

The measured environment variables are:

(1) Mother's participation in formal organizations (i.e., organizations of "Anglo" society), as an index of exposure to cultural materials, values, etc. of Anglo society. (Consists of 5 different measures.)

(2) Minority neighborhood vs. Anglo neighborhood.

(3) Cultural barriers, e.g., fluency of mother's English, knowledge of school, etc.

(4) Socioeconomic Status. An index based on the occupational status of the head of household and the total years of formal education.

(5) Urbanization--the extent to which child's parents were exposed to American urban society during their childhood.

(6) Home ownership.

(7) Individualistic achievement values--based on a composite of several questionnaires of values intended to assess the extent to which "Anglo" values had been internalized by the mother.

(8) Family structure, e.g., whether child lives with both biological parents, etc.

(9) Anxiety. (Sarason's School Anxiety Scale)--15 items assessed in an interview with child.

The multiple correlation (corrected for attenuation) between these 9 variables and Full Scale WISC IQ was 0.44 for Negroes and 0.37 for Whites, accounting for 19.6 and 13.6 percent of the IQ variance, respectively.

It is likely that these multiple correlations include more than just environmental variance. Some degree of genetic correlation is almost certainly involved in the multiple R between these environmental ratings and IQ if one acknowledges the compelling evidence for a genetic component in social class IQ differences within racial groups. So the multiple correlations of 0.44 and 0.37 in all probability considerably over-estimate the true correlation (i.e. independent of genotypes) between this set of environmental measures and the Wechsler IQ. (Also, remember that a multiple correlation to some extent capitalizes on chance, and when cross-validated in another sample the same regression coefficients will yield a somewhat lower R.) It is interesting that Mercer's R^2 (the proportion of variance attributable to ratings of the environment), even though it very probably contains genetic variance, does not exceed the proportion of environmental variance generally found in heritability studies of IQ based on twin and other kinship correlations, i.e. $1-h^2$, which is about 0.2 to 0.3.

Thus these environmental indices, which all together account for about 16% of the IQ variance (i.e., $R^2 = 0.4^2 = .16$) within racial groups, if applied to the between-groups variance (which in Mercer's samples is $[1.3\sigma/2]^2 = .42\sigma^2$) yields $0.16 \times 0.42\sigma^2 = .067\sigma^2$. On the IQ scale (with $\sigma = 15$), this is equivalent to about 4 IQ points out of the average 15 IQ points difference between the racial groups. (Mercer tries to explain the total IQ difference

by partialling out [i.e. statistically controlling] variables which are very likely highly correlated with parental genotypes for IQ [e.g., occupational status and education] and with race [e.g. living in a racially segregated neighborhood]). So we arrive again at the conclusion that environmental indices accounting for much if not all of the available environmental variance in IQ within racial groups, accounts for only about a third of the mean difference between the racial groups.

5. In the face of this analysis environmentalists must hypothesize the existence of as yet unidentified and unmeasured factors, which produce IQ differences between racial groups but do not contribute appreciably to IQ variance within groups. Since no one has clearly specified the nature of these factors, I shall simply label them "factor X." "Factor X" is purely ad hoc, invoked to explain the IQ gap still left when known, measurable environmental differences are taken into account. Notions such as "racial alienation," "white racism," consciousness of being a minority, identification with a historically mistreated minority, etc. are attempts to characterize factor X. While these factors may exist, it has not been shown, independently of the particular racial difference which they are invoked to explain, that they have any effect on IQ. And one may wonder why they do not apply to other minorities, such as Jews and Orientals, who also have been subjected to discrimination, etc., but who score at or above the national average on standard tests, or to American Indians, whose environmental deprivations are the most severe of any subgroup in the U. S. but whose performance on tests of mental ability and scholastic achievement is more or less intermediate as compared to whites and Negroes.

Psychometric Evidence

Although the discussion of racial differences from the standpoint of evolutionary theory and in terms of abstract principles of biometrical genetics can be carried on in general terms without reference to any particular racial groups, when we are faced with the prospect of actually making measurements and testing hypotheses we must get down to specific cases. At this point, understandably, there is often resistance or reluctance to our proceeding further. What may seem reasonable and intellectually acceptable in the abstract may seem odious and emotionally unacceptable when it comes down to specific cases.

It is a fact that the study of racial differences in mental abilities has focused much more extensively on sub-Saharan Africans and persons of African descent than on any other groups. Bibliographies of research on other racial groups are extremely scant by comparison. Because of the great technical and theoretical difficulties and uncertainties involved in the genuine cross-cultural testing of abilities, where language, customs, values, and the whole way of life differ markedly between the groups being compared, most investigators in differential psychology have chosen to study different racial groups which share a more or less common culture in terms of language, exposure to formal education, the forms of employment, and the cognitive demands associated therewith. The major racial groups in the United States, at least in recent decades, probably come closest to these criteria.

Numerically, Negroes are the largest of such racial minority groups in the U.S. population. In recent years a good part of the motivation for the psychological study of Negro-White differences in mental abilities has stemmed from the conspicuous and seemingly intractable differences in scholastic performance under fairly equal instructional conditions, and from the relatively

large percentage (more than three times that of whites) of Negro youths who fall below the minimum mental qualifications for induction into the armed forces, even when equated with the average white youth in amount of schooling.

This is naturally a sensitive subject, largely because of the history of racial discrimination in the United States and the Negroes' struggle to achieve equality of civil and political rights and opportunity for education and employment. Inferences about intelligence differences, whether measured by tests or manifested in scholastic and occupational performance, are viewed with dismay by many, also, I believe, because the vast majority of people correctly perceive what might be termed the "threshold" property of intelligence. That is to say, for many occupations in a technological society, there is some threshold or level of intelligence below which the ability to perform successfully is extremely improbable. There is a threshold of intelligence below which failure in school, as presently constituted, is virtually certain. And there is a threshold below which individuals are generally perceived as severely handicapped, socially as well as educationally and occupationally. Almost no other handicap--deafness, blindness, lameness, physical deformity--seems as generally overwhelming an impediment to achievement, self-realization, and what most persons think of as a satisfying life, as a very low level of general intelligence. We therefore naturally resist acknowledging evidence that a substantially larger proportion of some particular socially identifiable group, than of another group, falls below these various thresholds. The regrets and sympathies aroused by this perception, combined with feelings of guilt over deeply deplored historic injustices such as slavery and discrimination, has resulted in a common tendency in our thinking to transform these closely associated feelings into cause-and-effect rationalizations. It predisposes one to uncritical acceptance of explanations of certain racial differences in

cognitive abilities as due wholly to past or present social injustices, discrimination, poverty, and the exploitation of one people by another. Such concepts can be rich grist for political propaganda or goads to social action, but they are exceedingly remote from the kind of theoretical framework and fine-grained analyses of data that are needed if we really want to understand the existing evidence on particular racial differences in specific abilities.

Let us review briefly some of the main findings of psychometric research in the two groups that have been compared most extensively, viz., American Negroes and whites.

Magnitude of the Difference. Since mental abilities are seldom measured on an absolute scale, it is customary, for most tests, to describe the units of measurement in terms of the standard deviation of test scores in some representative sample of the population under study. Raw scores (i.e., number right) on mental tests called intelligence tests are usually converted to an IQ scale, with a mean of 100 and a standard deviation of 15 in the normative population.

White-Negro mean differences are most often expressed in units of the standard deviation within the normative population or within the white comparison group, which often amounts to about the same.

The magnitudes of the White-Negro test differences in all of the studies reported in the literature vary mainly in terms of several factors: age of the subjects, nature of the test, geographic region, and representativeness of the samples.

Age of Subjects. Tests devised for assessing the development of children under two years of age cannot be called intelligence tests, if by intelligence we mean the general factor common to performance on all complex cognitive tasks in the age groups above three or four years. Tests of whatever kind administered below two years of age show little or no correlation with cognitive tests

administered in later childhood and beyond. The infant tests, such as the Gesell, Griffiths, and Bayley scales, are reliable measures of early neuromuscular and sensory maturation and coordination. In the functions measured by those tests, Negro infants are considerably advanced as compared to white infants, up to 15 to 18 months of age. This infant precocity in motoric development has been noted also in a number of studies of African infants, as well as in Negroes in the U.S. In terms of a developmental quotient, with a mean of 100 and standard deviation of 15, such as provided by the Bayley scale, the White-Negro difference during these early months is of the order of 10 to 30 points. The largest differences on record favor African infants and U.S. Negroes in poverty areas in the South. This Negro precocity is also evident in physical indices of skeletal and neurological maturity at birth. There is also some evidence of Negro precocity in the earliest elements of language development, which is intimately related to motoric maturity. (Documentation of the research on all these points is given in Jensen, 1973a.)

By two years of age, the White-Negro developmental gap disappears. As the mental test content becomes more highly loaded with g (i.e., the general intelligence factor which accounts for most of the variance in complex cognitive tests in later childhood and maturity) with each succeeding year, the growth curve of the average white child overtakes that of the average Negro of the same age, and, by four to five years of age, the difference between the groups, provided the tests are highly g-loaded, amounts to about one standard deviation, equivalent to 15 points on the IQ scale, in favor of the white group. In g-loaded tests the White-Negro difference, expressed in standard deviation or σ units does not change after four or five years of age. I would speculate that this same difference of about 1 σ would be found as far down the age scale as the g factor can be measured. This hypothesis could be tested by comparing

the groups in terms of factor scores on the g factor rather than in terms of factorially complex test scores which have a diminishing g component as one moves down the age scale.

The fact that the Negro IQ deficit does not change at all beyond age five, relative to variation within either the White or Negro group, is of considerable theoretical importance. One of the main pillars of environmentalist explanations of the Negro IQ deficit is expressed by the so-called "cumulative deficit" hypothesis, which holds that environmental disadvantages act like compound interest in producing a cumulative deficit in Negroes' intellectual development. It has already been mentioned that Negro IQ declines from age 2, when it can first be measured, to age 4 or 5, after which it remains constant. This decline could be due to a cumulative deficit associated with certain environmental lacks, or it could be due to the increasing g loading of intelligence test items between 2 and 5 years of age. (By age 5 the g loading of intelligence tests like the Stanford-Binet already closely approaches its asymptotic value.) If the deficit were environmental, however, one must wonder why it does not continue to cumulate beyond age 5, when children enter school and are just becoming aware of the social milieu which environmentalists claim contain many of the key ingredients that depress Negro IQ and scholastic performance.

As important as the cumulative deficit hypothesis has been to the environmentalist program, I have not found any evidence to support it, and much evidence that contradicts it. Most studies of cumulative deficit have failed to control for possible demographic artifacts, such as differences in the populations sampled at various ages. But what is methodologically perhaps the most rigorous study of the subject, based on the IQ differences between younger and older siblings within the same families, using all the families in a California school

district with children between ages 5 and 12, and controlling for family size and birth order, there was found statistically significant evidence of a progressive deficit in verbal IQ requiring reading ability, but no evidence whatsoever of a cumulative deficit (as indicated by a zero difference between IQs of younger and older sibs) in a nonverbal, highly g loaded IQ test (Jensen, 1974a). Interestingly, the average White-Negro difference was at least as great on the Nonverbal as on the Verbal IQ test. The fact that the one standard deviation Negro deficit in nonverbal IQ is stable after age five, means that its causes, whatever they might be, must be sought in factors whose influences are already fully established before school age.

Nature of the Tests. The size of the White-Negro difference also depends upon certain properties of the test. Contrary to popular belief, verbal tests do not yield larger differences than nonverbal, and more often the reverse is true. However, my study of this matter leads me to believe that what little difference there is between Negro deficit in verbal and nonverbal tests is not in itself of fundamental significance. Verbal and nonverbal test batteries often reflect varying admixtures of two, more fundamental, classes of abilities, in one of which Negroes show little, if any, deficit, compared to Whites, and in the other of which Negroes show their greatest deficit (with the exception of one special ability, viz., spatial visualization). I call these two classes of ability Level I and Level II. Level I consists of abilities such as short-term retention of visual and auditory inputs, memory span, rote learning, and the like. It is characterized by reception, retention, and recall on cue, with a minimum of mental manipulation or transformation. Tests incorporating these features more or less exclusively can be made as demanding and difficult as one likes. They can require every bit as much of the subject's attention and effort as any other kind of test. We have used a

variety of such Level I tests in White and Negro samples and find little or no racial group difference relative to the individual variation within groups, which is considerable. Thus, an inteligence test that contains some items which can be acquired merely through familiarity, by repetition or rote learning, such as simple factual information and concrete vocabulary items, will to that extent reflect Level I ability. The Stanford-Binet and the Wechsler tests include some almost pure Level I tests, such as digit span memory. And to the extent a test is loaded with Level I, it minimizes the White-Negro difference.

Level II ability involves mental manipulation and transformation of inputs in order to arrive at a satisfactory output. This means discrimination, generalization, comparison, planned or goal-oriented search of immediately present stimuli or of stored memories, abstraction, classification, judgment, induction and deduction involving concepts. Level II is much the same as what Spearman termed g. The moment any mental manipulation, transformation, selection, or comparative judgment is aroused or demanded by the stimulus input, Level II or g enters the picture as a source of individual differences in the response. It is, of course, a greater source of variance the more the task calls for Level II processes relative to other sources of variance, such as Level I processes, sensorimotor abilities, attention, effort, and the like.

Test items that call for problem solving with novel materials, as contrasted with items that require recognition or recall of previously learned material, are the best measures of Level II, and they are the items with the highest g loadings when tests are subjected to factor analysis. Items such as those found in Raven's Progressive Matrices test are almost pure Level II, for example, while digit memory (i.e., repeating a string of digits immediately after hearing them spoken at a 1-sec. rate) is almost pure Level I. As soon

as we introduce some mental manipulation into the memory task, however, it takes on some Level II loading. It has been found, for example, that in a factor analysis of a number of Level I and Level II tests, forward digit span had nearly all of its factor loading on the Level I factor, while backward digit span (i.e., reciting the digit series in reverse of the order of presentation), had its factor loadings divided between the Level I and Level II factors, with slightly more on the latter. White and Negro groups differ most on the Level II factor and little, if at all, on Level I (Jensen, 1970, 1971, 1973c, 1974b).

A thorough survey of 382 studies involving some 80 different standardized intelligence tests on Whites and Negroes shows an average difference of about one standard deviation; the great majority of the group mean differences are between 10 and 20 IQ points (Shuey, 1966). All of these tests are predominantly g loaded, but many include other factors as well.

Attempts to show differences in the ability profiles of Whites and Negroes on tests of Verbal, Numerical, Figural Reasoning, and the like (e.g., Lesser, Fifer, & Clarke, 1965), I strongly suspect, are merely derivative, secondary phenomena reflecting the different Level I and Level II demands of the various tests. The available evidence does not appear to me to support the interpretation that Whites and Negroes have different profiles in the so-called Primary Mental Abilities themselves, except in so far as measures of these abilities cannot be divorced from their Level I and Level II demands. But there is one important exception, viz., spatial visualization ability.

A number of studies suggest that Negroes perform further below other groups (Whites, Orientals, American Indians, Eskimos) on tests of spatial visualization ability than on tests of any other ability. This has been found in Negroes of the West Indies as well as of the United States. The same tests

given to African Negroes show even lower scores, but they are not appreciably lower than a variety of g loaded tests which do not require spatial ability.

Spatial ability has long been suspected of being sex-linked, since it is the only one of Thurstone's seven Primary Mental Abilities which consistently shows an appreciable sex difference. Only about one-fourth of females exceed the male median in tests of spatial ability. Since Bock and Kolokowski (1973) have now demonstrated by quantitative genetic analysis that spatial ability is influenced by a single X-linked recessive gene, it is important from the genetic standpoint to see if this fact can help to explain the findings on spatial ability in American, West Indian, and African Negroes, and on the direction and relative magnitudes of sex differences in spatial ability, as compared with other abilities, in Negro and White groups. The present evidence, such as it is, appears consistent with the X-linkage of spatial ability and the additional fact that the 20 to 30 percent admixture of Caucasian genes in American Negroes came largely from male white ancestors, thereby resulting in the introduction of proportionally about one-third fewer X-linked than autosomal Caucasian genes into the American Negro gene pool. A rigorous test of this genetic hypothesis, however, awaits additional data (Jensen, in press). But it is of interest that quantitative genetics already has a theoretical model, in the mechanisms of X-linkage and recessivity, that appears capable of predicting the findings on White-U.S. Negro differences in spatial ability and their interaction with sex differences in spatial ability. Environmentalist explanations of these facts at present would have to be especially ad hoc, and would probably encounter difficulty with the fact that spatial ability, unlike g loaded tests, has relatively little correlation with socioeconomic status within racial groups.

Tests of scholastic achievement generally show slightly smaller White-Negro differences than most standard inteligence tests. This seems surprising

to many, but is consistent with the idea that some scholastic knowledge and and skills, such as spelling and mechanical arithmetic, are partly acquired by Level I processes. Scholastic tests which require the student to reason with his specific knowledge and skills to solve novel problems, however, are very highly correlated with general intelligence tests, and even with non-verbal tests of g, when all the testees have had the same number of years of schooling.

Geographical Region. The nationwide testing of youths for induction into the armed forces clearly reveals regional differences in intellectual ability, both for Whites and Negroes, though the regional differences are considerably larger for Negroes than for Whites. The White-Negro differences in various regions vary from the overall white average the equivalent of about 10 to 20 IQ points. Negro IQs are lowest in the South and Southeast and there is a gradient of increasing IQ as one moves further North and West. There is a similar, though less pronounced, gradient of IQ in the White population. This regional variation in IQ appears to be mostly a result of past selective migration associated with economic factors and employment opportunities making different educational and intellectual demands. It is of interest from our standpoint that variation in the amount of Caucasian admixture in American Negroes follows much the same regional gradient as IQ variation, from the Deep South, with close to 10% Caucasian admixture, to the North and West, with about 20 to 30%, and the Northwest as high as 40% (Reed, 1969). Since practically all the Caucasian genes in the American Negro gene pool were introduced during the period of slavery, which was confined to the South, the present regional variation is undoubtedly due to selective migration. It is significant that IQ and amount of Caucasian admixture in Negroes parallel one another in geographical distribution, and that both of these variables more or less parallel

the regional variations in the IQ in the White population.

Representativeness of the Sample. White-Negro comparisons have been reported where one or both groups are atypical samples of the White or Negro populations of a particular locality. Comparisons of White and Negro prisoners, juvenile delinquents, and patients in public hospitals, are examples. Such biased samples usually reduce the racial difference. The most frequent type of biased sampling is the matching of the racial groups on some index of socioeconomic status (SES), such as income and occupational and educational level. Such matching of the racial groups generally reduces their IQ difference by about one-third of a standard deviation, more or less, depending on how many IQ-correlated factors enter into the matching. It also depends, in the case of children, on whether one matches Negro and White children at the upper or at the lower end of the SES scale. High SES Negro and White children differ more in IQ than groups matched for low SES. In a review of the 33 studies before 1965, including a total of about 7900 Negro and 9300 White subjects, in which White and Negro groups were of comparable SES, Shuey (1966, p. 520) concluded: "The consistent and surprisingly large difference of 20.3 IQ points separating the high-status whites and high-status colored is accentuated by the finding that the mean of the latter groups is 2.6 below that of the low-status whites. It is probable that the home, neighborhood, and school environments of the white and colored lower-class children tested are more nearly alike in their stimulating qualities than are the home, neighborhood, and school environments of the white and colored upper and middle-class children; but it seems improbable that upper and middle-class colored children would have no more cultural opportunities provided them than white children of the lower and lowest class." Three more recent studies involving large samples also found low SES White children to have slightly higher IQs than middle and

upper SES Negro children (Wilson, 1967; Scarr-Salapatek, 1971; Jensen, 1974b).

The Hypothesis of Culture-Biased Tests

The most popular explanation of these psychometric differences, in whole or in part, is that the tests are in some way biased so as to favor Whites and disfavor Negroes. Since the tests often have been standardized on the white population, it is claimed that they are culturally loaded with content peculiar to Anglo middle-class experience, although this has certainly not been the intention of test constructors.

The claims of culture bias as an explanation of the White-Negro IQ difference in the United States runs into numerous difficulties. For one thing, many of the tests that show the greatest White-Negro difference show much smaller differences for other minority groups which are also regarded as disadvantaged or culturally different. On nonverbal IQ tests, which do not handicap children brought up in a foreign tongue, American Indians and Mexican-Americans outperform Negroes, on the average. On Raven's Progressive Matrices, one of the best highly g-loaded nonverbal intelligence tests, Arctic Eskimos. with their extremely different culture, score at least up to the White norms obtained in Scotland and the U.S. Chinese and Japanese in the United States at present score at least as high as native Whites, and in California they score higher, especially on highly g loaded nonverbal tests. Moreover, no one has yet devised or standardized an intelligence test within the Negro population which significantly narrows the racial IQ difference, although there have been serious attempts to do so. Yet most intelligence tests originating in the United States can be used in foreign countries simply by translating the test instructions and verbal items into the appropriate language. The translated tests retain highly similar reliability, validity, inter-item correlations,

and score distributions as are obtained in the U.S. white population. This has been the usual experience with the Stanford-Binet and Wechsler tests, which have been used in many countries with seldom more than translation and substitutions of a few of the informational items, such as changing "What is the population of the United States?" to "What is the population of Japan?" in the Wechsler test. A translation of the Stanford-Binet test into Negro ghetto dialect, however, produced no signficant increment (one IQ point, in fact) over the IQ obtained with the standard English version when given to Negro children most familiar with the ghetto dialect.

Recently, I have conducted intensive studies of culture bias in tests, using large samples of typical White, Negro, and Mexican children in California schools. I will here summarize the main results.

But first of all, one must distinguish between culture loading and culture bias. A test may contain informational content that could only be acquired within a particular culture. This can usually be determined simply by examination of the contents of the test items. Whether the particular cultural content causes the test to be biased with respect to the obtained scores between any two groups is a separate question. If the test includes only cultural content that is common to the experience of the groups being compared, it will not be culturally biased, assuming that the testing procedure itself is not a source of bias.

The fact that racial and social class groups differ on a test cannot itself be a proper criterion of bias. Legitimate criteria of test bias are of two types: external and internal. External bias is related to the predictive validity of the test, i.e., how well it predicts such criteria as school grades, success in some specialized training, and occupational performance. A test is biased if the intercepts and slope of the regression of criterion

measures on test scores differ significantly for the two or more populations in question. Reviews of the research on this point comparing White and Negro samples are unequivocal with respect to scholastic and job performance. There is a negligible difference in the slopes and intercepts of regression lines for Whites and Negroes. A single regression equation predicts equally well for both groups (Humphreys, 1973; Linn, 1973). Interestingly, the few exceptions reported in the literature would favor the Negro groups if the tests were used for selection, i.e., the difference in the regression lines is such that for any given test score Whites slightly out-perform Negroes on the criterion. In brief, the overwhelming evidence on the predictive validity of standard tests indicates that they are not biased against Negroes when compared with Whites. (There are too few studies of other ethnic groups to permit any general conclusions about them.)

It can, of course, be argued that the criterion predicted by the test scores is itself culture biased, and that one therefore needs a culture biased test to predict a culture biased criterion (e.g., scholastic achievement). Therefore, one must consider various internal criteria of test bias. These internal criteria seem especially appropriate for investigating the hypothesis that a given test is biased for one population when the item selection and standardization were based on a different population. If the test items are culture loaded, i.e., they call for specific information acquired in a particular culture, and if the cultures of the standardizations and target groups differ with respect to the cultural information sampled by the items, this should be reflected in the various internal indices of bias. I will list each of these indices and describe what we have found concerning each one with respect to White-Negro comparisons. (Mexican-American children were included also, but for the sake of simplicity I will not attempt here to summarize

these results. In general, they differ little from the results for Negroes, except that the Mexican subjects do relatively better on the nonverbal matrices test and relatively worse on the picture-vocabulary test.) All of these analyses have been made on what is probably the most culture loaded of all standard intelligence tests, the Peabody Picture Vocabulary Test (PPVT), and on one of the least culture loaded tests, Raven's Progressive Matrices.

The PPVT consists of 150 plates, each with four pictures. The examiner names one of the pictures and the subject is asked to point to it. The vocabulary ranges from very easy, common, and concrete words to very rare words and abstract concepts. The Progressive Matrices consists of 60 plates, each with a missing part which the subject must select from a multiple-choice set of six to correctly complete the pattern. Items range in complexity and difficulty from a level that is passable by most three-year-olds up to a level of difficulty beyond the capacity of the average adult. Figure 2 shows typical PPVT and Raven items of moderate difficulty.

- - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - -

The subjects in these studies numbered more than three thousand children in California schools, about equally divided among the racial groups. (I have presented these studies in detail elsewhere [Jensen, 1974c].)

1. Correlation of raw scores with chronological age in months does not differ appreciably for Whites and Negroes on either the PPVT or Matrices.

2. Internal consistency reliability and average inter-item correlation of both tests are the same for Negroes and Whites.

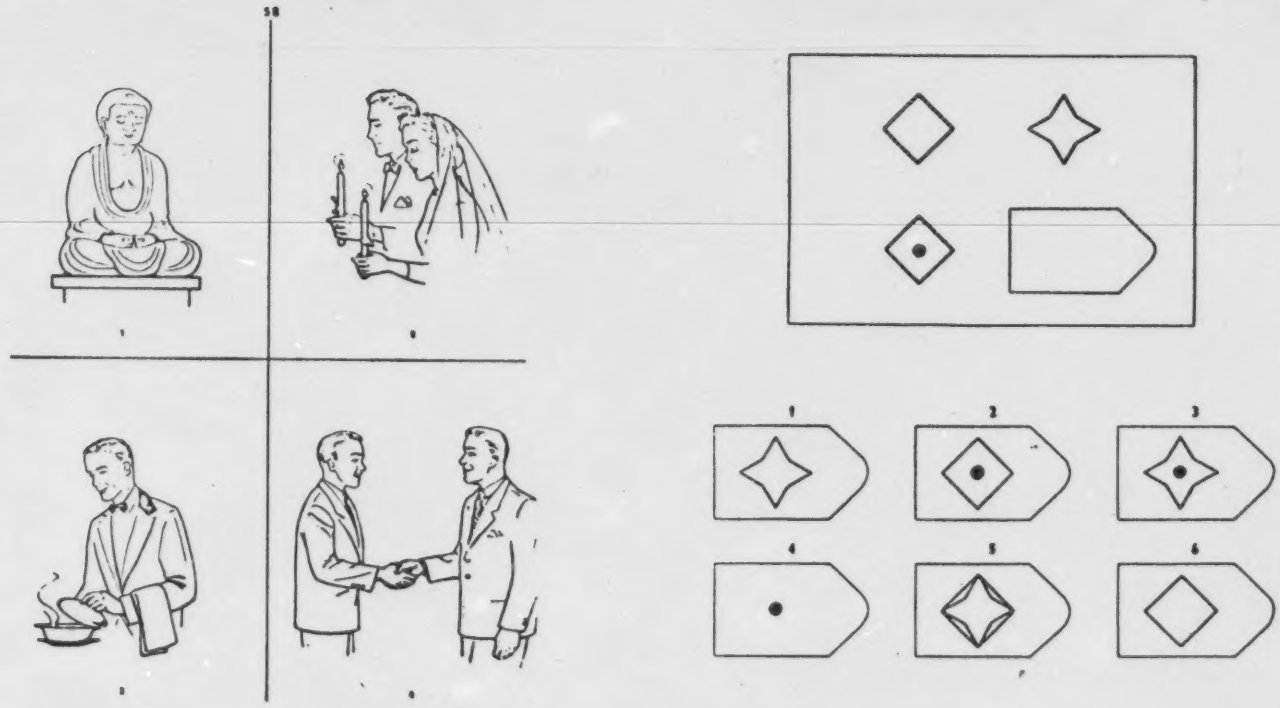

Fig. 2. Sample items of the Peabody Picture Vocabulary Test (left) and Raven's Progressive Matrices. The PPVT word for this item is "ceremony."

3. Rank order of item difficulty (as indicated by percent passing) is virtually the same in both racial groups. The correlation between $\underline{P}$ values (i.e., percent passing an item) over all items for Whites and Negroes is near perfect, without correction for attenuation. (For PPVT, $\underline{r}$ = .986; for Matrices, $\underline{r}$ = .993.) When the correlations are obtained for various subsets of 12 or 15 items, they are still very high (.87 to .99), and the correlations are highest in those subsets of items that discriminate most (i.e., have the largest differences in $\underline{P}$ values) between the races. This is the opposite to what one should predict from a culture bias hypothesis of the group differences, which should lead to the expectation that the most discriminating items would show the least similarity between the groups in the rank order of $\underline{P}$ values. In many subsets of items the correlation of $\underline{P}$ values between races is higher than between boys and girls within the same race, although boys and girls score about equally, overall. Certain PPVT items show more sex bias than any items show racial bias. For example, "parachute" versus "casserole" reflect different sexual biases in cultural knowledge. The PPVT also reveals culture biases in comparing white school children in England and white children in the United States. Although both groups obtain about the same total score, some vocabulary items are much easier for the English than for the Americans (e.g., "pedestrian" and "goblet,") and vice versa (e.g., "bronco" and "thermos"). Negro and white groups in California schools, on the other hand, do not show any of these marked discrepancies in order of item difficulty.

4. An even more sensitive index of cultural differences is the correlation between the item "$\underline{P}$ decrements" for the two races. The $\underline{P}$ decrement is the difference between the percent passing two adjacent items, e.g., $\underline{P}_1 - \underline{P}_2$, $\underline{P}_2 - \underline{P}_3$, etc. Thus we are measuring the racial group similarity in the differences in difficulty among items. Again, these correlations are very high

(.79 for PPVT, .98 for Matrices), and again the correlation was highest for the most discriminating sets of items. Correlations between the sexes within racial groups are not significantly greater.

5. Items that best discriminate individual differences within racial groups (i.e., items with the highest correlation with total test score) are the same items that discriminate most between the racial groups.

6. Incorrect responses (errors) are distributed in a non-chance fashion over the multiple-choice distractors in the same proportions for Whites and Negroes. There were several significant exceptions to this in the Matrices; that is, on some items Negroes made different errors than Whites. However, in every such instance it was found that the Negro children's proportions of responses to the various error distractors were the same as the proportions for white children who were approximately two years younger in chronological age. Thus it appears that the few differences that were found between White and Negro children are most clearly related to differences in level of mental maturity than to cultural differences.

7. The matrix of inter-item correlations for each test was factor analyzed within each racial group to determine the loadings of each item on the general factor (i.e., first principal component) that accounts for most of the covariance among all the items. The items' factor loadings for Negroes and Whites are highly correlated, and, most significantly, the correlation is markedly increased when the Negro factor loadings are correlated with the factor loadings of Whites who are about two years younger. In fact, on this index, 6th grade (ages 11-12) Negroes are more like 4th grade (ages 9-10) Whites than like 4th or 5th grade Negroes. (Also, 6th grade Negroes obtained about the same total raw score as 4th grade Whites.) Moreover, the loadings of items on the general factor within each racial group show a high positive

correlation with the degree to which the items discriminate between the races. In other words, those items which best measure what is common to all items within each race are the same items that show the largest race difference.

8. Few if any psychologists would claim that Raven's Matrices is more culture loaded than the PPVT. If the PPVT is culturally biased against Negroes, then, if we perfectly match PPVT and Matrices items for difficulty (i.e., percent passing) in the White population, we should expect, from the culture bias hypothesis, that these two sets of items would not be matched in difficulty in the Negro population. For Negroes, the culturally loaded PPVT items should be more difficult than the Matrices items. But, in fact, we found no significant difference. Thirty-five PPVT and Matrices items which are perfectly matched in difficulty for Whites turned out to be matched in difficulty for Negroes as well. (This was not true of Mexicans, for whom the PPVT items are significantly more difficult, as would be expected from the culture bias hypothesis.)

9. Finally, using an analysis of variance to examine the Race × Items interactions (for both PPVT and Matrices), we found we could almost perfectly simulate, without statistically significant differences, all features of the Negro-White differences, using entirely White samples. We simply divided the entire White sample into two groups, a younger group (ages 6 to 9) and a slightly overlapping older group (ages 8 to 11). Detailed comparisons of these two groups simulate, within the margin of sampling error, the results of the same comparisons of Whites and Negroes, when both groups are of the same chronological age. We have found no feature of the PPVT or of the Matrices which distinguishes Negroes from Whites who are about two years younger, or which distinguishes any differently between Negroes and Whites of the same age than between groups of younger and older Whites.

All these findings seem to me very incompatible with the culture bias hypothesis. To maintain this hypothesis one would have to postulate the additional and supremely ad hoc hypothesis that the cultural differences between Negroes and Whites perfectly simulate age differences within the White group, with respect to item difficulties, P decrements, inter-item correlations, choice of distractors, and g factor loadings, for tests as diverse as the PPVT and the Matrices.

In another study in which several mental tests were administered to several thousand White and Negro children by 12 White and 8 Negro examiners, it was shown that the race of the examiner had no significant or systematic effect on the intelligence test scores of White and Negro pupils (Jensen, 1974d). Also, special tests devised to measure attention, speed, persistence, and effort in the testing situation revealed only negligible differences between Negroes and Whites. I therefore conclude that these factors are an unlikely explanation of the large race difference in intelligence test scores.

Another study has shown that administering several mental tests under speeded conditions versus no time pressure did not significantly alter the White-Negro difference, although both groups performed better under the more lenient condition (Dubin, Osburn, & Winick, 1969). The same study also showed that pre-test practice on alternate forms of the tests did not significantly reduce the racial differences.

The most reasonable hypothesis, it seems to me, is that the two racial groups differ in the rate and the asymptote of development of the brain processes underlying the general factor common to intelligence test items. Comparisons of the racial groups across the ages from early childhood to adolescence on a number of different indices of mental growth lends further support to this hypothesis.

Consistency Among Developmental Indices

The Gesell Figure Copying Test (Ilg & Ames, 1964, pp. 63-129) consists of the ten geometric forms shown in Figure 3. The subject is encouraged simply to copy each figure, without time limit. A pencil with an eraser is an essential

- - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

part of the testing procedure. The test approximates a Guttman scale, i.e., it is like a series of hurdles, in that, if a subject can correctly copy, say, the fifth figure in the series, in all probability he can copy correctly all the preceding figures; and if he cannot correctly copy, say, the sixth figure, he will in all probability fail all the figures that follow it. The test reflects mental development over a range from about age 3 to age 12. When the Figure Copying Test has been factor analyzed along with standard intelligence tests, it is loaded almost entirely on the g factor.

We have given this test to more than ten thousand school children of different ethnic groups, of ages 5 to 12 years. There are marked group differences at every age, with Orientals scoring highest, followed closely by Whites, then Mexicans and, lastly, Negroes. The magnitude of the difference on this test is almost two standard deviations between Orientals and Negroes, as can be seen from Figure 4. Negro children in the 4th grade (ages 9-10)

- - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - -

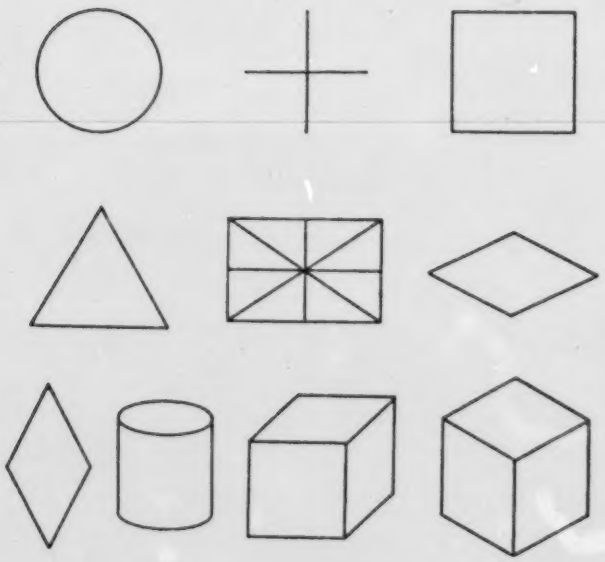

Fig. 3. The ten simple geometric forms used in the Figure Copying Test. In the actual test booklet each figure is presented singly in the top half of a 5-1/2" × 8-1/2" sheet. The circle is 1-3/4" in diameter.

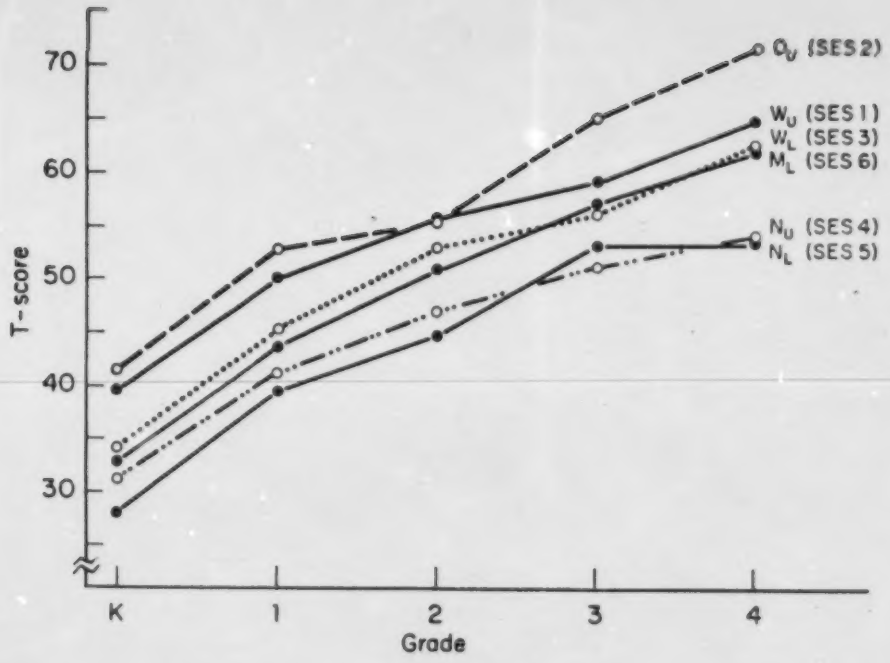

Fig. 4. Oriental (O), White (W), Mexican (M), and Negro (N) groups from socioeconomically urban, largely middle- to upper-middle class (U) and rural, largely lower- to middle-class (L) communities. The six groups are ranked from highest (SES 1) to lowest (SES 6) on a composite index of socioeconomic status.

perform on a par with Oriental children in the 1st grade and slightly below White children in the 2nd grade. The Mexican group, although lowest in socio-economic status, is almost exactly intermediate between the Orientals and Negroes and nearly on a par with Whites.

Even more telling is the fact that all these groups show the same developmental sequence of difficulties in copying these figures. The same conceptual difficulties appear in all the various ethnic groups, but simply at different ages, on the average. The difficulties of Negro children of ages 6 or 7 are indistinguishable from the difficulties of White and Oriental children of ages 5 or 6. Each figure, so to speak, "evolves," going from younger to older ages. Typical examples of some of the modal difficulties, as one goes from drawings of lesser to greater maturity, are shown in Figure 5. It would seem hard to explain in terms of cultural differences why Negroes

- - - - - - - - - - - - - -

Insert Figure 5 about here

- - - - - - - - - - - - - -

Whites, Orientals, and Mexicans all go through the same sequence of these peculiar characteristics of copying figures, and differ only in the average age at which they encounter the various difficulties up till the age at which they are able to copy the given figure correctly.

Jean Piaget has devised a number of highly diverse developmental tasks with similar properties. They are seemingly simple tasks, utilizing familiar objects, which call for judgment, mental manipulation, and reasoning about matters universally available to observation. One example is shown in Figure 6.

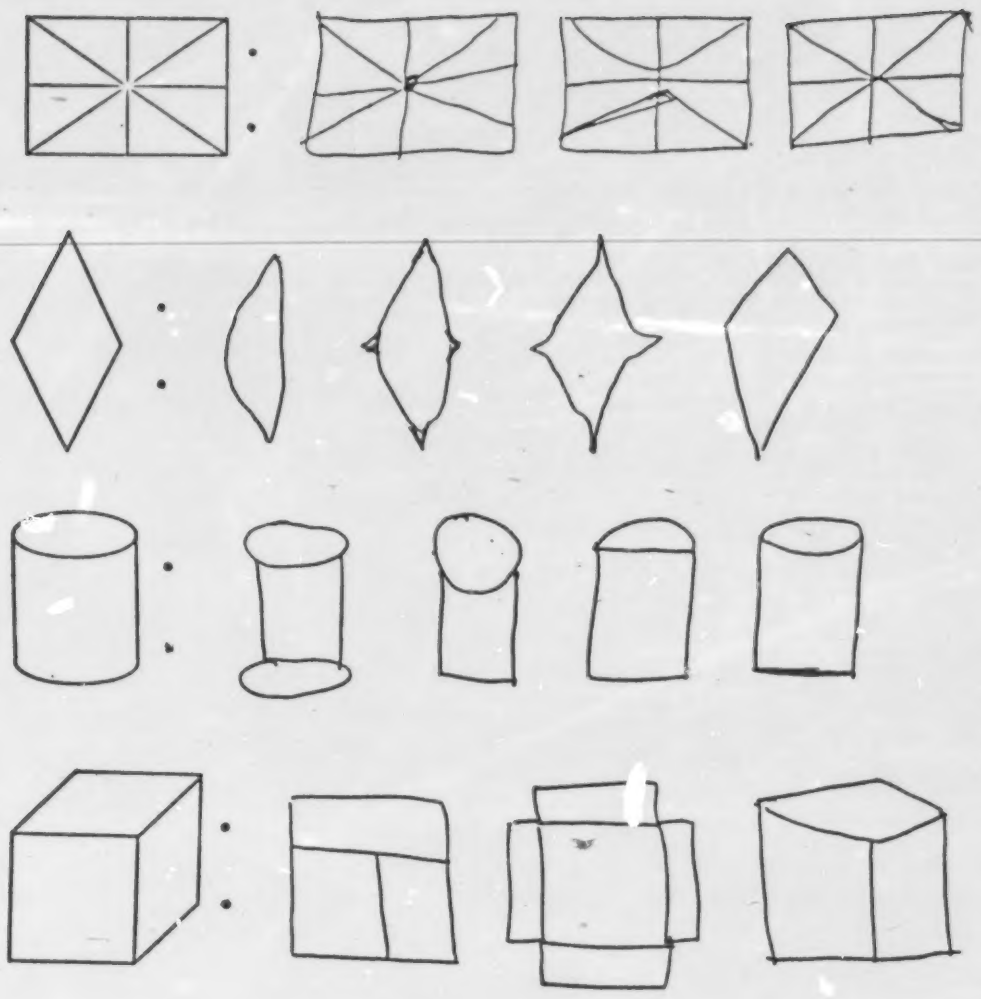

Fig. 5. Typical examples of developmental changes in children's figure copying, going from lesser to greater maturity. The models are on the extreme left.

- - - - - - - - - - - - - - -

Insert Figure 6 about here

- - - - - - - - - - - - - - -

The child is shown a bottle of red liquid (A). An opaque card is then placed in front of the bottle, and the bottle is tilted (B). Then the child is given a full-scale outline drawing of the bottle and is asked to draw the level of the red liquid as it will appear when the card is removed (C). Most children under 8 or 9 years of age draw a line more or less parallel to the bottom of the bottle, as shown in C, while older children more often correctly draw a horizontal line. Many children under 8 years do not markedly improve their drawing even after they have been shown the liquid in the tilted bottle. When this water-level test was given to large representative samples of three ethnic groups in Grades 1 to 3 (ages 6-8) in California schools, the percent of each group passing the test was Oriental, 43%; White, 35%; Negro, 13%. Tuddenham (1970) gave nine other such Piagetian tests of different concepts to the same groups. Negroes did less well than Whites on every item; Oriental children exceeded White children on 7 of the 10 items. The differences are comparable to those found with highly g loaded tests such as Raven's Matrices. (I have elsewhere reviewed in greater detail these and other studies showing similar results [Jensen, 1973a,pp. 312-318].)

Another developmental index is interesting because it has no right or wrong answers, but only preferences which change systematically with age. As children mature mentally, they show changing preferences for color, form, number, and size, in that sequence, in attending to the attributes of objects. The order of preference for children of kindergarten age (5 or 6 years) is

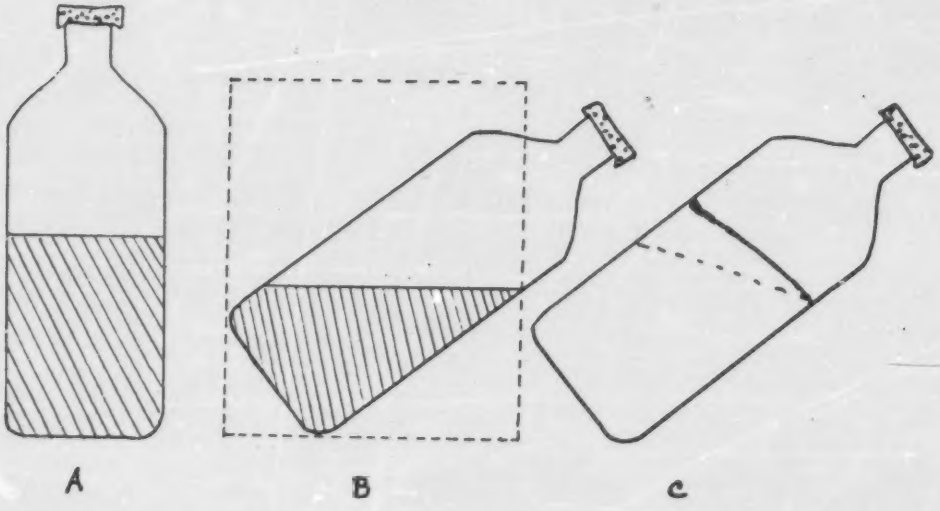

Fig. 6. Piaget's tilted water bottle test, to measure the concept of the horizontality of water level.

(1) form, (2) color, (3) number, and (4) size. Groups of White and Negro kindergarten children were each shown 12 different stimulus displays of the type shown in Figure 7.

- - - - - - - - - - - - -

Insert Figure 7 about here

- - - - - - - - - - - - -

The figures on the four cards differ simultaneously in color (green, red, blue, yellow), shape, size, and number. The examiner gives the small card at the top to the child and asks him to put it down on any one of the four cards with which he thinks it goes best. It is made clear that there is no "right" answer. Thus, the child can match the target card on the basis of color, form, number, or size. It turned out that White and Negro children of the same age differed in the relative frequencies of their preferences, in accord with the developmental prediction, i.e., the Negro children had a significantly higher percentage of the less mature preferences (Toki, 1971):

	Preference %			
	Color	Form	Number	Size
White	10.2	73.5	14.9	1.5
Negro	28.9	56.5	12.5	2.1

In view of the consistent racial differences in these highly diverse developmental tasks, I venture the hypothesis that typical Negro and White children will show consistent differences in growth rates and in orderly

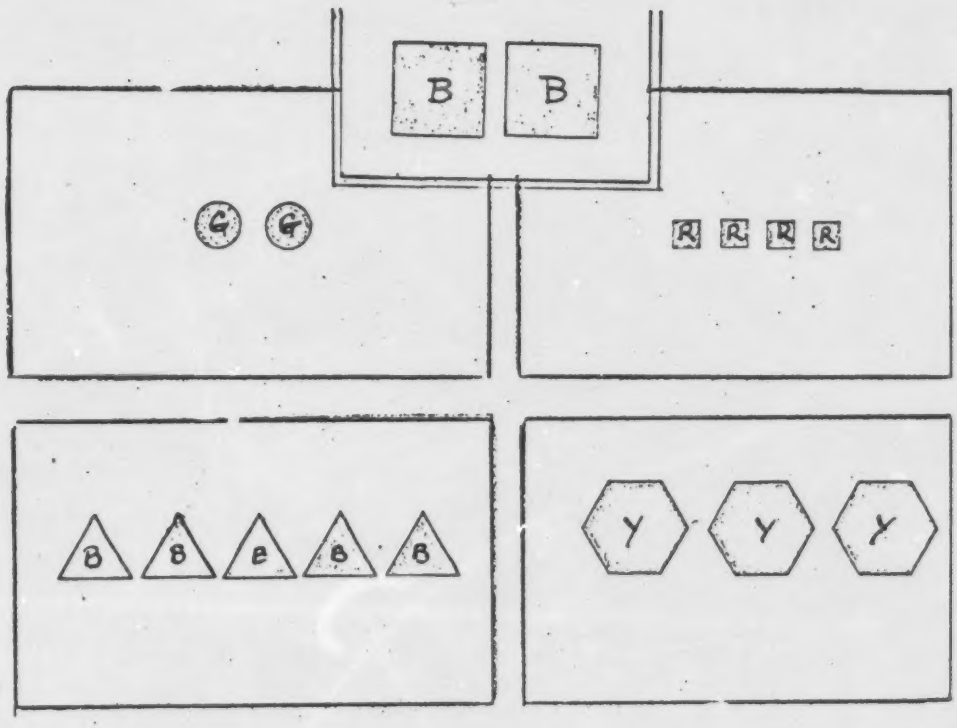

Fig. 7. An item of the color-form-number-size test. The letters (which do not appear on the actual test) indicate the colors of the figures.

sequential development on all age-related tasks involving abstraction, judgment, mental manipulation--in short, the essence of g or Level II ability.

Information Processing and the Essential Nature of g

Examination of the most highly g loaded test items shows them to be most clearly characterized by their requiring the testee to process a considerable amount of information--information not in the sense of merely recalling stored knowledge, but in the sense of having to take a number of facts and relationships into consideration simultaneously in order to produce or select the correct answer. Complexity, choice among alternatives, judgment, decision--these seem to be of the essence of g, as contrasted with memory, factual knowledge, and performance of highly practiced skills.

The idea of g as being related to information processing in this fundamental sense has come to the attention of a number of experimental psychologists, who have devised laboratory measures of information processing capacity. In order precisely to quantify the informational content of a task on an absolute scale, the task has had to be made very simple. And in order to demonstrate reliable differences in difficulty among tasks which, though they differ only slightly, differ by precisely known amounts in informational load, it is necessary to employ a very sensitive and continuous (rather than discrete) measure of the subject's response. The measurement of the subject's reaction time (in milliseconds) to stimuli meets this requirement. The complexity of the stimulus situation is varied so as to convey different amounts of information as measured in bits, the unit of measurement in information theory. A bit (for binary digit) is the amount of information that reduces uncertainty by 1/2. A stimulus situation to which the response is completely predetermined and therefore requires no discrimination, comparison, judgment, or decision has

no uncertainty and therefore conveys zero bits of information. The next most complex stimulus situation, involving two elements or alternatives, requires one decision and has one bit of information, e.g., Yes or No. Four alternatives have two bits of information, e.g., One or Two? Yes; One? Yes. Eight alternatives have three bits of information, and so on. The number of bits of information is the logarithm, to the base 2, of the number of alternatives. Very complex, highly g loaded test items undoubtedly contain many bits of information, but the actual number for any given item is not determinable by any means presently known. But using vastly simpler stimuli, though they are not nearly as good a measure of g because of their much smaller informational content, permits exact quantification of the task's complexity in terms of bits.

One of the important discoveries in this field, often called Hick's Law (Hick, 1952), which has been replicated in many studies, is that the subject's reaction time (RT) increases as a linear function of the amount of information as measured in bits. Thus, choice RT (i.e., responding differentially to two or more stimulus alternatives) is invariably greater than simple RT (i.e., response to a single stimulus). Hick's Law has been demonstrated by a variety of laboratory techniques, using different stimuli and different sensory modalities.

There are highly reliable individual differences in simple and in choice RT. Though it is not of much practical importance, it is of great theoretical significance that individual differences in simple RT are not significantly correlated with scores on standard intelligence tests, while choice RT is correlated with intelligence (Eysenck, 1967). The correlation is negative, i.e., the more intelligent subjects take less time to process a given amount of information.

We have devised an apparatus for very precise measurement of RT, as

well as movement time (MT), in response to stimulus arrays varying in informational content from 0 to 3 bits. This is an extremely small range of difficulty, so small, in fact, that persons have little or no subjective feeling that the 3 bits task is any more difficult than the 0 bit task. Even the experimenter cannot perceive a difference in the subject's RT in the 0 bit and 3 bits tests. But when precisely measured by electronic timers, RT increases regularly by some 30 to 50 milliseconds with each additional bit of information conveyed by the stimuli. These small fractions of a second, however, are subjectively negligible to subjects. This is very unlike ordinary intelligence tests in which the items often increase very perceptibly in complexity and difficulty, even to the point that the increasing appearance of difficulty can possibly intimidate subjects and discourage continuing effort.

The apparatus for measuring the subject's RT and MT consists of a panel, 13" × 17", painted flat black, and tilted at a 30° angle. At the lower center of the panel is a red pushbutton, 1/2" in diameter, called the "home" botton. Arranged in a semi-circle above the "home" button are eight red pushbuttons, all equidistant (6") from the "home" button. Half an inch above each botton (except the "home" button) is a 1/2" faceted green light. Different flat black panels can be fastened over the whole array so as to expose arrays having either 1, 2, 4, 6, or 8 light-button combinations.

The subject is instructed to place the index finger (of his preferred hand) on the "home" button. Then an auditory "ready" signal is sounded (a high-pitched tone of 1 sec. duration), followed, after a continuous random interval of from 1 to 3 seconds, by one of the green lights going "on," which the subject must turn off as quickly as possible by touching the sensitive microswitch button directly under it. RT is the time the subject takes to remove his finger from the "home" button after the green light goes on. MT

(movement time) is the interval between removing the finger from the "home" button and touching the button which turns off the green light. RT and MT on each trial are registered in milliseconds by two electronic timers. On each trial never more than one light in the whole array goes "on," and the subject turns it off by touching the button adjacent to the light. The particular light that goes "on" in each trial is completely random and thus is unpredictable by the subject, thereby creating the uncertainty upon which the quantification of information depends.

Our experiments with this apparatus have shown, in accord with Hick's Law, that RT increases as a perfectly linear function of bits of information, in school children and in young adults. The linear increase in RT as a function of bits shows up for individuals as well as for the group as a whole, and is therefore a very lawful and reliable psychological phenomenon. The average correlation between RT and bits for individual subjects is over 0.9, which means that even for individuals (and not just for the group average) there is an almost perfect linear regression of RT on bits of information. MT, on the other hand, is completely unrelated to bits and remains constant across all amounts of information, both for individuals and for the group means. The reliabilities of both RT and MT are about .90 when Ss are given 30 trials on each light/button combination. Also, it is apparent that RT and MT are not measuring the same sources of individual differences variance, since the correlation between the two measures is only about 0.3. Moreover, there is virtually no functional relationship between RT and MT, as indicated by a within-subjects correlation between RT and MT of close to zero.

A measure of information processing capacity that is independent of absolute RT is the slope, b, of the linear regression of RT on bits. When the regression of RT on bits is determined for every subject, the values correlate

significantly but lowly (around -0.30) with standard intelligence test scores. This accords with the hypothesis, as described by Eysenck (1967), that information processing capacity, assessed independently of absolute RT in terms of the rate of increase in RT as a function of the increasing complexity of the task, is a significant component of intelligence.

Now what has all this to do with our topic of race differences? In a recent study we hypothesized the following: If the essence of the g component in the subject's performance is related to the degree of complexity of the task (in the information theory sense), and if Whites and Negroes differ in g capacity, then Whites and Negroes should show no significant difference in performance on tasks of zero information (e.g., one light/button combination or simple RT), but should show increasing differences as the number of bits of information increases, even when the increases in information are all within such a narrow range and at such a low level of complexity (i.e., between 0 and 3 bits) as to be subjectively indiscriminable in difficulty.

We tested this hypothesis on 200 male youths, 18 to 19 years of age, with nearly equal numbers of Negroes and Whites. The samples were not representative of the general population. All subjects were within the normal range of intelligence and the White and Negro groups were almost perfectly matched in the score distributions on a group verbal test of general intelligence. They were also matched as closely as possible in years of schooling (averaging 11.5 years), although the Negroes averaged about half a year more schooling than the Whites. Many studies have matched racial groups on socioeconomic, educational and other environmental factors correlated with intelligence test scores, and have shown that the group difference is diminished by such matching, often with the claim that if more such environmental factors had been controlled, the test difference would be wiped out completely. The

present racial samples go a step further: they do not differ appreciably in test scores. This stacks the cards against our hypothesis that the groups would differ increasingly in an information processing measure as the amount of information increases. Negro and White groups which score the same on more or less culturally and educationally loaded verbal paper-and-pencil tests, if measured on a series of much less culturally and educationally loaded tests involving regularly increasing amounts of information corresponding to g, should show regularly increasing differences as the amount of information increases. And this should be shown to occur even within such a narrow and easy range of cognitive difficulty as not to be subjectively perceptible by the subjects, so that attitudinal and motivational factors would be a most unlikely cause of any differences. Also, there is no *a priori* reason to believe that such factors would affect RT for different amounts of information independently of MT (movement time). We already knew from previous studies that information processing was related to RT but not to MT, which is uncorrelated with the informational content of the task.

The results turn out perfectly in accord with the hypothesis. The White and Negro groups differed negligibly and nonsignificantly (about 3 milliseconds) in mean and median RT to 0 *bits* (i.e., the one light/button task), but the White-Negro differences in RT increased significantly and linearly with each additional *bit*, at the rate of 10 milliseconds per *bit*. At three *bits* the groups differ 31 milliseconds in RT, a highly significant difference ($p < .001$). The slope of the regression of RT on *bits* was determined for each individual, and the difference between the mean slopes of the White and Negro groups was shown by a *t* test to be significant ($p < .01$), with the Negro group showing the steeper slope, i.e., greater increments in RT as the *bits* of information increase. (Note that the slope measure is independent of absolute RT.) The increase of RT as a function of *bits* shows no

significant departure from linearity in either group. Slope also correlated significantly (about 0.30, $p < .01$) with the mental test scores within groups. Also, it is interesting that intra-individual variability (i.e., an individual's variation about his own mean over repeated trials) increases systematically as a function of bits. The rate of this increase was very significantly greater for the Negroes; in fact, it was the largest racial difference to show up in any of the measures derived from this testing procedure.

The groups also differed significantly on MT (White < Negro), but MT showed no correlation with bits.

The multiple correlation, R, between several of the RT and MT measures, on the one hand, and the racial dichotomy, on the other, is 0.41.

The important and indisputable point of this study is that the two racial samples, which differ much less (in fact, not significantly) in ordinary psychometric scores than do the general populations of Whites and Negroes, and in which the Negroes have more education than the Whites, still show highly significant differences in a behavioral task, in accord with prior expectations based on theoretical considerations of the essential nature of g and information processing capacity.

These findings, however, are not without precedent. In unselected samples of Negroes and Whites, even larger differences have been found in choice RT. Noble (1969) found a highly significant ($p < .001$) difference between Negro and White children (matched for age and sex) on a 4-choice RT test. In a sensitive measure of speed of visual information processing (requiring no motor response at all), using a visual recognition test involving only 2 bits of information (i.e., 4 alternatives), Bosco (1970) also found a highly significant difference between a group composed mostly of low SES Negro children and a group of middle SES White children. Poortinga (1972)

measured simple and choice RT to both auditory and visual stimuli in groups of native African and European students in South Africa. (The groups differed 2.89σ on Raven's Matrices.) Choice RT (2 and 3 bits) for both the visual and auditory stimuli showed the Africans to have significantly longer mean RTs; in units of the White group's standard deviation, the White-African difference was 1.9σ for auditory and 1.5σ for visual RT. But there was no difference between the groups in simple RT (i.e., 0 bits.) In terms of our hypothesis, however, some doubt is raised about the interpretation of these striking results by the fact that visual and auditory choice RT showed negative though nonsignificant correlations with Raven's Matrices in the African sample and a significant correlation (-0.45) only for auditory choice RT in the White sample. Such puzzles, of course, simply indicate the need for further experimental analysis.

Environmentalist Hypotheses

Environmental hypotheses of IQ differences have been largely ad hoc: each hypothesis that falls down under rigorous scrutiny is immediately replaced by a new one, which enjoys popularity until investigators have tried, but fail to find supporting evidence. Scientifically, all evidence is not of equivalent weight. Ad hoc evidence is less impressive than theoretically predicted evidence. A theoretical model which accommodates existing facts and predicts new facts is vastly preferable to an assemblage of ad hoc criticisms and explanations.

Innumerable hypothesis have been put forward to explain the White-Negro difference strictly in environmental terms. Those which have been formulated clearly enough to be tested have not stood up when put to the test. I have reviewed the claims of environmentalists about each of numerous factors

said to explain the lower Negro IQ--inequality of schooling, teacher expectancy, motivation, language deprivation, nutrition, and reproductive casualty. None of them adequately accounts for the facts of Negro performance on mental tests (Jensen, 1973a, 1973b). As researchers find each of the more obvious environmental factors, such as those associated with SES, not to hold up as explanations of the Negro IQ deficit, more subtle, often unmeasurable, environmental influences have been hypothesized. In the past few years, each newly proposed environmental hypothesis has failed as soon as it was put to the test.

One of the most popular of recent hypotheses has been that the majority of Negro children have a different language than standard American English, and this supposedly handicaps them in scholastic performance and in taking IQ tests. But how would this explain the results on nonverbal tests? And why do immigrant children with little or no knowledge of English not show a similar deficit? Why do children who were born deaf and are therefore severely language-deprived and score lower on verbal tests show no deficit on nonverbal tests? A recent comprehensive review of the research evidence pertaining to the "different language" hypothesis found no support for it--"In general, no acceptable, replicated research has found that the dialect spoken by black children presents them with unique problems in comprehending standard English" (Hall & Turner, 1974, p. 79). The investigators believe the explanation of the Negro deficit must be sought elsewhere, stating that they " . . . are convinced that more effort should be directed toward studying universals of cognitive development rather than toward relatively superficial performance differences such as spoken dialects" (p. 80).

Another popular environmental hypothesis is that the cause of the Negro IQ deficit is to be found in the quality of the mother-child interaction during the preschool years. The hypothesis is admittedly difficult to test, since

appropriate investigation must rely upon naturalistic, systematic, comparative observations of Negro and white children in their natural psychological environments. Two developmental psychologists, Alfred and Clara Baldwin (1973), have spent more than a decade conducting this kind of investigation, including several hundred records of mother-child interactions involving pre-schoolers in both Negro and white families from lower and middle social class. Many aspects of mother-child interaction (35 coded variables) were systematically observed and recorded in half-hour long free-play settings. Only one significant ethnic difference showed up: Negro mothers were more likely than white mothers to adopt a didactic teaching role in free play. The Baldwins note that " . . . white mothers were much more relaxed in general about the child's academic future. They felt considerably less pressure to teach him academic-type facts during the play session than did the black mothers" (Baldwin & Baldwin, 1973, p. 720). They continue: "On no other measures did we find ethnic differences. The amount of interaction was not consistently different for the black and white groups; the level of syntactic complexity was not different if educational level is held constant. Except for the fact that didactic teaching does involve more direct behavior requests, we saw no evidence that black mothers were more bossy or more punitive. In fact, we observed very little punitiveness in any of the play sessions." In the light of their observations, the Baldwins believe the language deprivation theory is called into question: "All these facts lead us to question deeply whether there is any social significance in the small difference in the syntactic complexity found in the mothers in the free-play session." They admit, "Frankly, when we began this investigation, we anticipated many more differences between the black lower-class sample and the white upper-middle-class sample. . . . But as we observed these mother-child pairs, and then as we saw the results

of the data analysis, we have become convinced that the most striking fact is the overall similarity of mother-child interaction in free play in all the samples" (p. 720).

In view of the failure of numerous environmental hypotheses to be borne out by evidence, the genetic hypothesis appears reasonable and highly likely, which is not to say that it is proven. But at least it is already established that genetic factors are the most important determinant of IQ differences within the racial groups, and, in the absence of any compelling environmental explanation for the White-Negro intelligence difference, we would be scientifically remiss not to seriously consider the genetic hypothesis.

In terms of what is already known about human evolution, about a host of other kinds of genetic racial differences, about the relative contributions of genetic and environmental factors to differences in mental abilities, and about the constancy (relative to the variability within groups) of White-Negro differences in IQ and a wide variety of other indices of cognitive development from childhood to maturity, it appears highly probable that genetic factors are involved to a substantial degree in the lower average IQ of American Negroes. So far, I have not seen a serious attempt to adduce evidence, or comprehensive argumentation based thereon, to the effect that this hypothesis is either improbable or scientifically unwarranted.

I have focused on differences between Whites and Negroes in the U.S. only for illustrative purposes and because there has been vastly more relevant research on representative samples of these populations than is true of any other racial groups. I have little doubt that other racial populations can be shown to differ behaviorally in complex ways, both in the cognitive and personality domains, and it would seem most surprising if genetic as well as cultural and environmental factors were not involved in many of these

differences.

Finally, it seems scarcely warranted in this context to emphasize the points that races are not Platonic absolutes, that races cannot be ordered on an overall continuum of "superior-inferior," since human differences are multidimensional, that differences are biologically and socially important only in relation to particular environmental demands, and that human racial differences *per se* have no obvious, direct, or value-free implications for social policy.

References

Baldwin, A. L., & Baldwin, Clara P. The study of mother-child interaction. American Scientist, 1973, 61, 714-721.

Baker, J. R. Race. New York: Oxford University Press, 1974.

Bock, R. D., & Kolakowski, D. Further evidence of sex-linked major-gene influence on human spatial visualizing ability. American Journal of Human Genetics, 1973, 25, 1-14.

Bosco, J. J. Social class and the processing of visual information. Final Report, Project No. 9-E-041, Contract No. OEG-5-9-325041-0034(010). Office of Education, U. S. Dept. of Health, Education, and Welfare, May 1970.

Burt, C. Quantitative genetics in psychology. British Journal of Mathematical and Statistical Psychology, 1971, 24, 1-21.

Burt, C., & Howard, M. The multifactorial theory of inheritance and its application to intelligence. British Journal of Statistical Psychology, 1956, 9, 95-131.

De Fries, J. C. Quantitative aspects of genetics and environment in the determination of behavior. In Lee Ehrman, G. S. Omenn, & W. E. Caspari (Eds.), Genetics, environment and behavior: Implications for educational policy. New York: Academic Press, 1972. Pp. 5-16.

Dubin, J. A., Osburn, H., & Winick, D. M. Speed and practice: Effects on Negro and White test performances. Journal of Applied Psychology, 1969, 53, 19-23.

Erlenmeyer-Kimling, L., and Jarvik, L. F. Genetics and intelligence: a review. Science, 1963, 142, 1477-1479.

Hall, V. C., & Turner, R. R. The validity of the "different language explanation" for poor scholastic performance by black students. Review of Educational Research, 1974, 44, 69-81

Hick, W. E. On the rate of gain of information. Quarterly Journal of Experimental Psychology, 1952, 4, 11-26.

Humphreys, L. G. Implications of group differences for test interpretation. Assessment in a Pluralistic Society. Proceedings of the 1972 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1973. Pp. 56-71.

Ilg, Frances, L., & Ames, Louise Bates. School readiness: Behavior tests used at the Gesell Institute. New York: Harper & Row, 1964.

Jensen, A. R. Hierarchical theories of mental ability. In Dockrell, B. (Ed.) On intelligence. Toronto: Ontario Institute for Studies in Education, 1970. Pp. 119-190.

Jensen, A. R. A two-factor theory of familial mental retardation. Human Genetics (Proceedings of the Fourth International Congress of Human Genetics). Amsterdam: Excerpta Medica, 1971.

Jensen, A. R. Educability and Group Differences. New York: Harper & Row, 1973.a (London: Methuen, 1973).

Jensen, A R. The differences are real. Psychology Today, 1973, 7, 80-86. b

Jensen, A. R. Level I and Level II abilities in three ethnic groups. American Educational Research Journal, 1973, 10, 263-276. c

Jensen, A. R. Cumulative deficit: a testable hypothesis? Developmental Psychology, 1974. a

Jensen, A. R. Interaction of Level I and Level II abilities with race and socioeconomic status. Journal of Educational Psychology, 1974, 66, 99-111. b

Jensen, A. R. How biased are culture-loaded tests? Genetic Psychology Monographs, 1974, 00, 000-000. c

Jensen, A. R. The effect of race of examiner on the mental test scores of White and Black pupils. Journal of Educational Measurement, 1974, 11, 1-14. d

Jensen, A. R. A theoretical note on sex linkage and race differences in spatial visualization ability. Behavior Genetics, in press.

Jinks, J. L., & Eaves, L. J. IQ and inequality. Nature, March 22, 1974.

Jinks, J. L. & Fulker, D. W. Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. Psychological Bulletin, 1970, 73, 311-349.

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monographs, Society for Research in Child Development, 1965, 30, No. 4.

Linn, R. L. Fair test use in selection. Review of Educational Research, 1973, 43, 139-161.

Lorenz, K. Die Rückseite des Spiegels: Versuch einer Naturgeschichte menschlichen Erkennens. Munich: R. Piper Verlag, 1973.

Lush, J. L. Genetic unknowns and animal breeding a century after Mendel. Transactions of the Kansas Academy of Science, 1968, 71, 309-314.

McClearn, G. E., & De Fries, J. C. Introduction to Behavioral Genetics. San Francisco: W. H. Freeman, 1973.

Mercer, J. & Brown, W. C. Racial differences in IQ: Fact or artifact? In C. Senna (Ed.) The Fallacy of IQ. New York: The Third Press, 1973. Pp. 56-113.

Nei, M., & Roychoudhury, A. K. Gene differences and effective divergence times between the three major races of man: Caucasoid, Negroid, and Mongoloid. Genetics (Proceedings of the 13th International Congress of Genetics), 1973, 74 (No. 2, Part 2), 193. Abstract.

Noble, C. E. Race, reality, and experimental psychology. Perspectives in biology and medicine, 1969, 13, 10-30.

Poortinga, Y. A comparison of African and European students in simple auditory and visual tasks. In L. J. Cronbach & P. J. D. Drenth (Eds.) Mental Tests and Cultural Adaptation. The Hague: Mouton, 1972. Pp. 349-354.

Reed, T. E. Caucasian genes in American Negroes. Science, 1969, 165, 762-768.

Scarr-Salapatek, Sandra. Race, social class, and IQ. Science, 1971, 174, 1285-1295.

Shuey, Audrey M. The Testing of Negro Intelligence. (2nd ed.) New York: Social Science Press, 1966.

Toki, Ellengale. Kindergarten children's preferences for color, form, size, and number stimuli. Unpublished master's thesis, University of California, Berkeley, 1971.

Tuddenham, R. D. A 'Piagetian' test of cognitive development. In Dockrell, B. (Ed.) On intelligence. Toronto: Ontario Institute for Studies in Education, 1970. Pp. 49-70.

Urbach, P. Progress and degeneration in the IQ debate. British Journal of the Philosophy of Science, 1974 (Summer).

Wilson, A. B. Educational consequences of segregation in a California community. In Racial isolation in the public schools, Appendices, Vol. 2 of a report by the U. S. Commission on Civil Rights. Washington, D. C.: U. S. Government Printing Office, 1967.

Footnote

[1]De Fries (1972, pp. 10-11) states that "Unfortunately, no valid estimate of r is available." But then he goes on to suggest a value of r based on a coefficient of inbreeding. The value of r is approximately twice the coefficient of inbreeding. He uses a coefficient of inbreeding estimated from morbidity data in Hawaii to arrive at a value of r of .002. This is the average intraclass correlation (among different racial groups in Hawaii) for a random sample of all gene loci. There are many genes, perhaps the vast majority, that have not been subject to selection and which have similar frequencies in all human populations. Gene frequencies would differ only from random drift for most of the genes that enter into a coefficient of inbreeding estimated from morbidity statistics. Such an average over all loci does not provide any clue as to the intraclass genetic correlation for polygenic traits that have been subject to selection pressures as intelligence undoubtedly has. The intraclass genetic correlation for skin color in Europeans and Africans, for example, would be much higher than .002 and probably approaches 1.00. The same would be true of height in Pygmies and Watusis. To what extent this is true for intelligence, we do not know. Obviously, the De Fries formula can yield no estimate of h^2_B unless we can obtain as estimate of r for the specific polygenic trait in question. An estimate of r based on the average correlation over all loci, or on a random sample of genes, or on some other trait simply will not do, and to base speculations on such estimates can only be misleading.